2015

Designing a Data-Tracking System for a Private Therapeutic Day School

Olin J. Bittner
Antioch University - Seattle

DESIGNING A DATA-TRACKING SYSTEM

FOR A PRIVATE THERAPEUTIC DAY SCHOOL

A Dissertation

Presented to the Faculty of

Antioch University Seattle

Seattle, WA

In Partial Fulfillment

of the Requirements of the Degree

Doctor of Psychology

By

Olin J. Bittner

October 2015

DESIGNING A DATA-TRACKING SYSTEM

FOR A PRIVATE THERAPEUTIC DAY SCHOOL

This dissertation, by Olin J. Bittner, has
been approved by the committee members signed below
who recommend that it be accepted by the faculty of the
Antioch University Seattle at Seattle, WA in partial fulfillment
of requirements for the degree of

DOCTOR OF PSYCHOLOGY

Dissertation Committee:

_____
Bill Heusler, Psy.D.
Chairperson

_____
Steven Curtis, Ph.D.

_____
Shamsah Ebrahim, Ph.D.

_____
Date

iii

ABSTRACT

DESIGNING A DATA-TRACKING SYSTEM

FOR A PRIVATE THERAPEUTIC DAY SCHOOL

OLIN J. BITTNER

Antioch University Seattle

Seattle, WA

The Children's Institute on Mercer Island (CHILD) is a private therapeutic day school in

the Seattle area serving students in elementary and secondary education. Their stated

mission is to "provide innovative school programs and therapies that promote social,

emotional and academic development for children with special needs." In the fall of 2012

they engaged in a program evaluation that in many respects resembles a needs assessment

in order to explore and improve aspects of their functioning. Through preliminary

evaluation processes, including dialogue with CHILD's Leadership Team and a survey of

internal stakeholders, an area of interest in student mental health was uncovered and an

initial evaluation question emerged: "How does CHILD claim expertise, particularly in

the area of mental health?" Historically, evidence of CHILD's impact in this regard has

been largely anecdotal. Aside from a limited collection of behavioral data pertaining

largely to IEP goals and objectives, CHILD does not track mental health, or long-term

student outcomes. As a program interested in its own claims to "expertise," members of

the Leadership Team and other stakeholders have called for improved data collection in

this regard. This program evaluation is an attempt to understand the types of data that

would be most useful to CHILD's interest in expertise and then design a program for

tracking this data. Bronfenbrenner's (1994) ecological model was used as a framework

for guiding data collection, the results of which are synthesized and integrated into a series of recommendations constituting the final results of the project. This dissertation is available in open access at AURA, http://aura.antioch.edu/ and Ohio Link ETD Center, https://etd.ohiolink.edu/etd

Table of Contents   Page

This dissertation is dedicated to the staff and students of CHILD.

It is also dedicated to my wife, Rachel, who puts up with me.

List of Figures   Page

Introduction

Children's Institute of Learning Differences on Mercer Island (CHILD) is a

private therapeutic day school in the Seattle area serving students in elementary and

secondary education. In the fall of 2012 they initiated a program evaluation to explore

and improve aspects of their functioning. Through preliminary evaluation processes,

including dialogue with CHILD's Leadership Team and a survey of internal stakeholders,

an area of interest in student mental health was uncovered. Historically, evidence of

CHILD's impact in this regard has been largely anecdotal. Aside from a limited

collection of behavioral data pertaining largely to IEP goals and objectives, CHILD has

not systematically tracked mental health, or long-term student outcomes. This is in spite

of the fact that, over the course of the past 40–50 years, there has been a steady rise in the

utilization of evidence-base practice (EBP) in education. In this climate, schools, school

districts, and other educational programs have been placed under increasing pressure to

demonstrate for others the effectiveness or impact of their activities on student outcomes.

Accountability in education has become paramount. As a substrate of that system, special

education has been subject to the same forces, and finds itself today increasingly

concerned with EBP (Cook & Cook, 2013) and the need, in general, to demonstrate

positive outcomes. While the particulars of EBP may vary across domains, EBP in

general tends to connote a collection of practices shown to be effective through research.

Moreover, this tends to incorporate some form of quantitative data collection and

analysis, often times in ways that fit both with broader policy and the more local needs of

the schools, their students, and the related communities (Cook & Cook, 2013; Forness,

2005; McDuffie & Scruggs, 2008). While EBP is not the focus of the present evaluation,

its prevalence in the field is referenced here to illustrate the importance of tracking outcomes empirically.

In light of these developments within the field of special education, CHILD has taken an interest in the outcomes, or evidence-base of its own practices. The preliminary phase of this project, described in greater detail in the Methods section, was geared towards the development of the evaluation question, a central aspect of program evaluation (Fitzpatrick, Sanders, & Worthen, 2011). Through conversations with CHILD's Leadership Team and a survey of internal stakeholders, the central question to emerge from this preliminary process was "How does CHILD claim its expertise, particularly in the area of mental health?" Operating as one of several private therapeutic day schools in the Seattle metropolitan area, CHILD's mission is to "provide innovative school programs and therapies that promote social, emotional and academic development for children with special needs" (Children's Institute for Learning Differences [CHILD], 2014, About Us section, para. 1). This is in keeping with assertions made by leaders in the field (Greenberg et al., 2003) that "school-based prevention programming—based on coordinated social, emotional, and academic learning—should be fundamental to preschool through high school education" (p. 467). They have little, however, in the way of reliable and valid data demonstrating student outcomes in the areas of social and emotional, or mental health. Not only would data of this nature provide evidence of program impact or efficacy, it could also help CHILD differentiate itself from other schools of its kind. Up to this point, the primary and perhaps only post-placement outcome data has been in the form of anecdotal reports from the school's primary consumers: parents and school districts. The school has no formal data collection system

designed specifically to track student mental health progress and longitudinal outcomes.

Therefore, in light of recent trends in evidence-based practice and mental health in

schools, the aim of this project is to first evaluate what sorts of data might be most useful

to the school, then to develop for them a formalized system with which this data might be

collected, stored, analyzed, and used. The ultimate focus of the evaluation, in other

words, sought to answer the question, "How can CHILD demonstrate effectiveness, and

therefore claim expertise in the area of mental health?"

As a way of organizing and structuring the data collection and analysis phases of

the project, Bronfenbrenner's (1994) ecological model will be used as a template. In this

model, individuals or organizations are understood to be influenced or defined by

interaction at multiple levels: the micro-level, or immediate environment of a given

entity, the meso-level, or the system of relevant organizations and communities, and the

macro-level, or the broader political and cultural terrain. The assumption here is that

proper understanding of a given entity requires analysis at all three of these levels (Doll,

Spies, & Champion, 2012). CHILD's ecological system will be defined and explored for

the purposes of determining the potential data most relevant to tracking student mental

health outcomes. Once these data points, or indicators, have been established, a plan for

tracking them will be developed and articulated in a series of formal recommendations.

Background and Literature Review

To better understand the sociocultural dynamics that have in part necessitated this project, it's important to provide historical context. Thus, the first section of the literature review traces the development of federal policy in education and the impact this has had on special education at both the national and local levels. Broader trends in policy and programming have also given rise to an interest in mental health in the school system. An initial review of the literature and research in this area is necessary to ensure the proposed data collection system makes use of and is appropriately situated within the most current developments in the field. This proposal will review the relevant research in the area, highlighting its impact on the field and those areas in need of further exploration and analysis. While the process of evaluation is often fluid and emergent, those areas most salient at this juncture of the project will be explored and a base for future literature review drawn out. The specific plan for the evaluation will be detailed in the Methods section.

Policy in Historical Context: Implications for Special Education Programs

In an attempt to address the growing achievement gap between students in different demographic regions of the country, federal legislation was passed in the form of the Elementary and Secondary Education Act (ESEA) of 1965. This bill instituted a new era of accountability for schools by requiring states to set and regulate achievement standards in exchange for federal funding. Originally consisting of six "titles," its "Title I" made ESEA the first bill to authorize spending for the education of children with disabilities (originally termed "educationally disadvantaged," funding from this provision

was reappropriated towards "educationally deprived" children and schools in 2002). The ESEA has been re-authorized a number of times, most recently in 2001 as the No Child Left Behind Act (NCLB), which tightened accountability measures by emphasizing standardized assessments and local control of schools (Whilden, 2012). As Fitzpatrick et al. (2011) point out, this shift to a culture of accountability was reflected in such documents as the report published in 1983 under the Reagan Administration entitled *A Nation at Risk*. The report's message, according to Fitzpatrick et al., was that the federal government was needed to step in and fix a "broken" system, from which a "federal role with a focus on accountability emerged" (p. 54). The ESEA, as federal legislation, remains the largest overhaul of the nation's education system in history. Today, 90% of public and parochial schools nationwide continue to receive funding under provisions laid out by the law (Brown-Nagin, 2012).

While students with special needs were recognized in early versions of the ESEA, there were no specific protections in place to guard against discrimination in the school system. Students with special needs were often excluded from public schools entirely, and parents were left alone to find a solution for their child's academic needs. The landscape began to shift, however, with two landmark legal cases in 1972: Pennsylvania Association for Retarded Children (PARC) v. Commonwealth of Pennsylvania and Mills v. Board of Education of the District of Columbia.

The PARC case pertained directly to the rights of school districts and schools to exclude students based on their disability. Up to that point in Pennsylvania, children who had not attained a "mental age" of five by the age of eight were considered unlikely to "profit" from public school and could thus be legally excluded. This ruling was

challenged and ultimately found to be unconstitutional (PARC v. Commonwealth of Pennsylvania, 1972), paving the way for future legislation that would establish every child's right to a free and appropriate public education, or FAPE (Simon, 2005). In Mills v. Board of Education of the District of Columbia, also brought to court in 1972, it was determined that financial considerations could not be prioritized over a child's right to an education, regardless of cost. By 1975, federal statutes had been passed establishing unprecedented legal protections for special needs students. The Education for All Handicapped Children Act (EAHCA) of 1975 included the concepts of least restrictive environment (LRE) and individualized education plan (IEP). Both of these mandates remain in place today, ensuring students' rights to optimal placement and service while holding schools and school districts accountable to their progress. The EAHCA was updated again in 1990 and renamed to the Individuals with Disabilities Education Act, or IDEA, which continues as the legal cornerstone to our present-day special education system.

A core component of the ESEA and its subsequent revisions has been concerned with the distribution of federal tax dollars and the way this funding is utilized in the education system. Funneled down to the schools and school districts on a state-by-state basis, it is deliberately linked with student achievement. When NCLB's predecessor, the ESEA, was passed in 1975, it established the standard of adequate yearly progress (AYP)—a standard that continues to operate under the NCLB—requiring states to track statewide student progress through standardized academic tests. If states are interested in retaining federal funding, their students must maintain a state-determined, federally-approved AYP. In other words, states must demonstrate continued progress across time to

receive their share of federal education dollars. According to Cook and Cook (2013), this process of accountability has, over time, led to increased demand for measurable evidence of program and school efficacy. With the NCLB revisions in 2001, standardized testing became an annual procedure between the third and eighth grades for all but a small portion (1%) of the students in each state. This change left many states scrambling to come up with the money to simply pay for the testing, and has led to increased scrutiny of and controversy around the federal legislation. While many states have successfully sought waivers to ease the financial and systemic burden this has imposed, there remains a prevailing theme of accountability as it relates to student progress (Cohen, 2006).

Mental Health in Schools

At the same time that federal policy has become increasingly stringent in its accountability practices, an interest in the way mental health issues are being addressed in the school system appears to have emerged. According to Levitt, Saka, Romanelli, and Hoagwood (2007), this is due in part to the "high rates of unidentified and untreated youth with mental health problems" (p. 164) and has led to a call for the widespread adoption of such practices as universal mental health screening. In 1995, the Maternal and Child Health Bureau's (MCHB) Office of Adolescent Health in the U.S. Department of Health and Human Services created an initiative entitled *Mental Health of School-Age Children and Youth*. The initiative designated grant funding for two areas related to student mental health needs: improving the infrastructure of mental health services within the school system, and creating innovative resources and instructional materials for use in a school setting. One outcome of this initiative was the creation of two national, university-based resource centers: the Center for Mental Health in Schools at UCLA, and

the Center for School Mental Health Assistance at the University of Maryland (Adelman et al., 1999). According to the Center for Mental Health in Schools (n.d.), its focus is dedicated to "promoting healthy development and addressing barriers to learning at a school site in ways that can expand the impact of mental health in schools" (para. 10). The sheer volume of research and information at both sites gives some evidence of the emerging importance of mental health in the school system.

This expansion of interest has been reflected in federal policy-making. In 2003, former President George W. Bush commissioned the President's New Freedom Commission on Mental Health (2003), a year-long task force dedicated to the study of mental health needs and care in America. Among its findings, it was recommended that

Federal, State, and local child-serving agencies fully recognize and address the mental health needs of youth in the education system. They can work collaboratively with families to develop, evaluate, and disseminate effective approaches for providing mental health services and supports to youth in schools along a critical continuum of care. (p. 62)

In their summary, the Commission cited a growing research base indicating positive outcomes for school-based mental health programs. Of particular note was their role in improved student engagement and test scores, and decreases in incidents involving disciplinary action. The institution of universal screening for mental health problems was also recommended (Levitt et al., 2007).

The President's New Freedom Commission's report says that "while schools are primarily concerned with education, mental health is essential to learning as well as to social and emotional development" (2003, p. 58). Recommendation 4.2 of the report says to "Improve and expand school mental health programs" (p. 62). This reflects ongoing efforts at the both the grassroots, professional level/s as well as the policy and legal levels to more effectively integrate education and mental health. Not only does this report

illustrate the increasing relevance of mental health in the education system, it substantiates CHILD's interest in demonstrating positive mental health outcomes amongst its students. The following repositories and models, while not directly pertaining to CHILD, are examples of the way in which mental health is being addressed within the national school system.

   CASEL: Collaborative for Academic, Social, and Emotional Learning. In 1994, a group of educators and philanthropists founded the Collaborative for Academic, Social, and Emotional Learning (CASEL) to "support schools and families in their efforts to educate knowledgeable, responsible, and caring young people who will become productive workers and contributing citizens in the 21st century" (Elias et al., 1997, p. viii). It is worth noting up front that CASEL tends to use the phrase "social and emotional" in place of "mental health" in their literature, in contrast to other documents, such as the New Freedom Commission discussed above which use "mental health" unsparingly. Part of the reason for such inconsistencies may lie in the fact that federal special education legislation under IDEA uses "mental health" largely in reference to mental health providers outside of the education system, as in efforts should be made amongst educators to connect students to such providers who are out in the community. The phrase "mental health" appears only once in relation to provisions of services within the education system. According to the Individuals with Disabilities Education Act (2002), federal funds may be used to "assist local educational agencies in providing . . . appropriate mental health services for children with disabilities" (Part B, 611 (C)(iii)). Perhaps as a result of this legislation, and as noted below in the Methods section, school districts may regard "mental health" services as a discreet type of service due to the

financial arrangements associated with the use of the phrase. Members of CHILD's Leadership Team have said they avoid its use with school district representatives because they do not necessarily want to imply a given student is in need of specific services above and beyond what is typically offered in their program, services that would incur an additional expense to the school district. CHILD's Leadership Team requested that the phrase "social and emotional" be used in place of "mental health" for these reasons when interacting with school district representatives. While "social and emotional" does not appear at all in IDEA, it may be understood as a suitable synonym for mental health due to programs such as CASEL and their prevalence in the special education research and literature.

According to Elias (2004), "the term 'social-emotional learning' (SEL) was developed for use in research and practice . . . as applied to the schools because it reflected a strong recognition of the role of both social and emotional factors in successful academic learning" (p. 54). It is not a defined approach to intervention, in other words, but a term used to encapsulate and organize a class of ideas. Walberg, Zins, and Weissberg (2004) identified a number of common characteristics or themes in their review of SEL-related research. These included the self-management or regulation of stress, goal-setting and problem-solving skills, socially engaging teaching strategies, and caring and collaborative working relationships, amongst others. However, while the Collaborative for Academic, Social, and Emotional Learning (2013) recently released their 2013 CASEL Guide to evidence-based SEL programs, they include no set criteria defining social and emotional learning. Instead, it appears to exist as a growing body of related, if somewhat loosely defined, research and practices. For instance, in a meta-

analysis of 213 school-based social and emotional learning (SEL) programs involving 270,034 K–12 students, Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) found that participants in SEL programming demonstrated significantly improved social and emotional skills, attitudes, behavior, and academic performance. Studies included in the analysis fit broadly into categories related to school-based social and emotional interventions, and the dependent variables consisted of six different student outcomes: social and emotional skills, attitudes toward self and others, positive social behaviors, conduct problems, emotional distress, and academic performance. While they include in their analysis studies assessing for mental health symptoms such as anxiety and depression, the authors generally avoid the use of formal diagnostic language. Social and emotional learning in their view incorporates a wide array of psychological and academic functioning. This overt omission of the phrase "mental health" is indicative of the topic's political nature, and while an analysis of this sort is beyond the scope of this project, it provides a backdrop for the discussion. "Social and emotional" and "mental health" will be used somewhat interchangeably in this project, with the exception of interactions with school district representatives, as described above.

One example of a formalized school-based, SEL intervention is PATHS (Providing Alternative THinking Strategies), a prevention-based program focused on "the development of essential developmental skills in emotional literacy, positive peer relations, and problem solving" (Greenberg, Kusché, & Riggs, 2004, p. 172). Behaviorally at-risk youth in a special education setting were administered a PATHS curriculum in a randomized control trial (n = 49 in the experimental condition; n = 59 in the control group) experiment. Using cognitive and academic achievement measures,

it was found that mathematics ability improved along with spatial problem-solving and processing speed. Social-cognitive and behavioral indicators also showed an improvement in emotional understanding, feeling identification, conflict resolution and affect regulation relative to the control group. These findings underscore the importance of considering mental health, or social and emotional functioning, in educational decision- and policy-making and programming.

    PBIS: A school-based, psychological-oriented approach. One other area of development worth mentioning in mental health-related school intervention and practice is the emergence of Positive Behavior Interventions and Supports, or PBIS. PBIS is the only type of support or intervention referenced in the 1997 re-authorization of the Individuals with Disabilities in Education Act (IDEA), and as such, IEP teams are actually required by law to consider the use of PBIS in their determination of student needs and services (OSEP Center on Positive Behavioral Interventions & Supports: Effective Schoolwide Interventions, 2013). PBIS emphasizes data-driven decision-making and the organization of resources for the improvement of model fidelity through "an implementation framework that is designed to enhance academic and social behavior outcomes for all students" (Sugai & Simonsen, 2012, p. 1). As a stipulation of federal mandate, PBIS carries special weight in the consideration of school-based mental health practices. It also connotes a specific approach to school-based mental health intervention that emphasizes functional behavioral assessment and behaviorally-oriented intervention (Positive Behavioral Intervention and Supports, 2015). While it's unclear if CHILD currently incorporates PBIS into their programming, it may be necessary to incorporate PBIS principles into the structure of the recommended data-tracking system.

In summary, CHILD's interest in tracking mental health outcomes is a function of a broader movement to demonstrate progress or impact empirically. Their question, "How can CHILD demonstrate effectiveness, and therefore claim expertise in the area of mental health?" arises out of these macro-level cultural dynamics, and serves their interests at a meso- and micro-level. They would like to be able to demonstrate empirically to external stakeholders such as parents and school district representatives the impact they are having on the mental health of their students. As noted above, mental health has become an entrenched focus of stakeholders at all levels of the education ecosystem, and CHILD's interest to this end arises as a part of their core mission "to provide innovative school programs and therapies that promote *social, emotional* and academic development for children with special needs" [emphasis added].

CHILD: An Overview

To effectively utilize information collected from the literature, it's necessary to understand how and to what end CHILD intends to operate in the first place. This can be achieved in what is referred to in the evaluation literature as program theory development, or the process by which the evaluator, through conversations, observations, and reviews of documents and relevant research, develops a model of how a program is theorized or designed to work. This is often done quite literally in the form of a logic model, or a visual map or flowchart of intended program resources, operations, and outcomes (Fitzpatrick et al. 2011). A logic model has been developed for this project, and is located in Appendix A. The section that follows explains much of the information contained in the logic model, and provides a written overview of CHILD, including its operations, goals, and model of intervention. Many aspects of both the logic model and

program overview will remain the same throughout the project, but the logic model itself is a changing document that will be increasingly oriented towards the project's objectives. As new information is uncovered and aspects of the program—either real or intended—are identified, the logic model will be refined in relation to the project's purpose of designing a data-tracking system. All of the information contained in this section has been provided by members of CHILD's Leadership Team, except where otherwise noted.

CHILD is a private, regional, therapeutic day school and clinic located in Mercer Island, WA. It serves students from kindergarten through 12th grade presenting with a variety of learning disabilities, mental health issues, and sensory processing or neurological conditions. It currently serves students from 19 school districts across six counties, although these numbers may fluctuate slightly throughout the year due to changing enrollment. At its inception in 1977, it served one school district, but has grown to serve more than 20 from six different counties across its 36 year history. CHILD is classified both as a non-public agency (NPA) and as an approved independent school in the state of Washington. It is a fee-for-service program that contracts with the school districts to provide services to students whose needs cannot be met through their regularly assigned schools. CHILD's status as an NPA requires oversight from the Washington State Office of the Superintendent of Public Instruction (OSPI) in the form of annual approval processes and a tri-annual site visit and audit. It is one of a number of programs in the area working with similar school-aged populations, and is accredited through the Northwest Accreditation Commission (NWAC).

As an NPA, CHILD also has considerable autonomy to operate in the way it sees fit. However, the contracts they have with school districts stipulate that students must be making progress while at CHILD (due in part to AYP mandates discussed earlier). This progress is measured primarily through each student's individualized education plan (IEP), a legal document that defines the scope and nature of special education services needed. The documents are generally comprised of a variety of assessment documents and school records, and outline specific goals for the student. Quantifiable progress on these goals acts as the primary determinant in the evaluation of a student. CHILD has some influence in the process, and can adjust IEPs as necessary provided they are acting in accordance with the relevant state and federal statutes. This allows for a degree of flexibility in the system, and probably increases the chances a given student can make measurable progress since goals can be changed readily to meet individualized needs. However, as described above, this undermines the rigor of the collected IEP data, and renders it unreliable as a source of objective evidence of progress. Hence, CHILD endeavors to establish a more robust data-tracking system, one that can be used to demonstrate expertise in the area of mental health through reliable and valid data collection—assuming the data indicates student progress while enrolled at CHILD. A very real possibility is that analysis of the data does not show adequate or sizable mental health progress. In this case, the data may be less useful as evidence of expertise, and more useful as an internal feedback tool to improve services. This consideration could factor into this project's final proposal, insofar as CHILD is interested in developing a data-tracking system with multiple objectives. In a more general sense, the Leadership Team at CHILD expresses concern that mental health issues are often ignored or de-

prioritized within the special education system. This may be a justified concern given that one in five students experience mental health issues in the K–12 population (President's New Freedom Commission on Mental Health, 2003), Their interest in a data-tracking system is thus an attempt, in a more general sense, to draw awareness to and provide support in the area of student mental health.

Program structure. There are seven classrooms at CHILD, each with a lead teacher and two to four instructional assistants. There are generally between five and eight students per classroom. A second tier of staff known as "prevention" provide support to the classrooms and play a critical role in working with children in crisis. There are four members on the prevention team, each one assigned to two different classrooms, although they support each other depending on the needs of the students and staff throughout the day. All prevention staff are trained in the Pro-Act model of intervention, which trains staff to de-escalate and calmly intervene when a child is escalated, while minimizing threats and harm to self and others. All classroom and therapy staff are trained in Pro-Act course-work and have the opportunity to observe modeling by the prevention staff. CHILD does not use restraints or seclusion rooms, but if necessary the appropriately trained staff can intervene physically to deescalate a dangerous or excessively disruptive situation. However, this is viewed as a last resort; CHILD takes some pride in not being a "hands-on" school.

In addition to the teachers and prevention staff, there is a team of specialists that includes occupational therapists (OTs), speech and language pathologists (SLPs), counselors, an art specialist, and an in-house videography specialist. The OTs, SLPs, and counselors are a part of the Developmental Therapy Services (DTS) department at

CHILD, which serves children at CHILD and in the community. The art and videography specialists work creatively with students to make art projects and movies. The clinical director in charge of these programs is also one of four members of CHILD's Leadership Team. The other three are the executive director, the director of education, and the director of community relations. A small number of other staff members round out the remaining administrative, clerical, and janitorial positions. Finally, a volunteer board of trustees offers additional leadership in support of policy and governance issues.

Student profile. Enrollment at CHILD fluctuates, but is generally around 45 students. Maintaining a consistent enrollment is important because a large portion of their budget is dependent on student tuition. Students may be referred by the student's school district, or by parents of the students, who pay privately for the services. Scholarships or breaks in tuition are sometimes awarded for private pay clients, but generally the tuition is around $60,000 a year per student. The students, who stay two to three years on average, are in need of services above and beyond what is available to them in their public school special education programs. Each student's IEP team is required to place the student in the "least restrictive environment" in which they can be expected to succeed. For those at CHILD, a determination has been made that their needs would be best met in a separate school environment. Presenting mental health issues range from language and communication disorders, autism spectrum disorders, ADHD, trauma and crisis, to anxiety disorders, sensory processing disorders, and school avoidance. Many suffer from multiple disabilities. In the multi-tiered response-to-intervention (RtI) framework, CHILD serves what is considered the tertiary level of the overall student population, or the top 5% in terms of need.

According to Wholey (1996), a crucial aspect of program evaluation is helping administrators and managers choose appropriate performance or outcome measures as well as the tools with which to measure them. When the goals or desired results of the program are clearly articulated, it becomes easier to understand what types of data might best indicate whether the goals have been achieved. According to their director of education, CHILD's general goal is to return its students to their usual public school special education classrooms within one to three years of their arrival at CHILD, better prepared to navigate and respond to the social and performance demands of school. In more formal terms, their mission is "to provide innovative school programs and therapies that promote social, emotional and academic development for children with special needs." CHILD also utilizes a number of other orienting strategies that both inform and are informed by their value structure. Recently, a number of staff convened to discuss the core values and purpose of CHILD, a meeting that resulted in a short document titled "The CHILD Way." This details in brief one to two line statements the basic approach and duties of the CHILD program. A main objective for the students (outside successful IEP progress, which is the school's contractual obligation), as outlined in this document, is to develop what CHILD refers to as their "three 'R's": regulation (ability to self-regulate), relationships (e.g., social skills), and resilience. CHILD's goals for its students will determine in part the most appropriate types of data to track.

Collaborative Problem Solving: An intervention model. The primary model of intervention used by all staff at CHILD is Dr. Ross Greene's Collaborative Problem Solving (CPS) approach (Greene, 1998). In this model, children's difficulties are understood to be a function of their own "unsolved problems" or "lagging skills," an

attempt on Greene's part to dispel commonly-held beliefs that problematic behavior is a failure of motivation and/or a function of manipulative, attention-seeking behavior. Rather than a top-down approach relying on token economies and behavioral reward/consequence interventions, it encourages staff to seek out the underlying causes of the different behaviors of its students with the understanding that "kids do well if they can." If the staff understands the student's problem, the theory goes, a collaborative interaction ensues in which student needs and teacher expectations are both met (Greene, 2011). This model was developed as a cognitive-behavioral intervention for children and adolescents with aggressive behavior and has served in part as an alternative to seclusion- and restraint-based models of intervention (Greene, Ablon, & Martin, 2006). A study by Martin, Krieg, Esposito, Stubbe, and Cardona (2008) found that implementing CPS in an inpatient setting significantly reduced the frequency and duration of both seclusions (when an escalated client is placed in an isolated and secure room) and restraints. Greene et al. (2004) also demonstrated that CPS was somewhat more effective than a parent-training intervention in treating youth with oppositional defiant disorder. This model has been used more frequently in treatment programs than in school settings, but its evidence base indicates possible efficacy in this arena.

CPS was adopted by CHILD in 2005 as an alternative to their former model, which relied in part on restraints and seclusions. Recently, CHILD has taken steps to ensure fidelity of model implementation. In May of 2013, its founder, Dr. Greene, visited the school for an all-day seminar with the staff, a training event the Leadership Team plans to reinforce through monthly training sessions and seminars via videoconferencing technology. Furthermore, discussions with the CHILD Leadership Team, as well as

results from an informal stakeholder survey did not reveal any major concerns over program fidelity. Issues of model fidelity may be of concern down the road, especially if CHILD begins tracking student outcomes more rigorously. In the meantime, however, this project, based on conversations with CHILD's Leadership Team and a survey of internal stakeholders (described below), assumes they are taking steps to solidify their practice in this regard. The more important aspect for the project at this point is in the role CPS plays as a mental health intervention. Apart from CHILD's mental health specialists, the teachers are trained in education, not necessarily in mental health. Therefore, mental health outcomes may be related to the effectiveness of the CPS model. As a data tracking system is developed, it may be helpful or necessary to understand the nature of this relationship. For instance, should the number of staff trainings in CPS be included in the tracked data set? Or, if CPS is designed to teach students certain skills, might these skills be measured as one possible mental health outcome? These questions and others will be considered in light of the relevant literature and survey results.

Methods

In many ways, this evaluation is taking a case study approach. According to Fitzpatrick, et al. (2011), "case studies are often used for descriptive purposes when the desire is to examine an issue from many different perspectives" (p. 392). CHILD and its Leadership Team are interested in understanding how CHILD can claim expertise in its service delivery model, particularly in the area of mental health. However, as suggested above, exploration at different levels is required, both within the school and without. What is the cultural/political climate such that this question is necessary, for instance? How does CHILD fit into this broader cultural terrain? How is mental health viewed within the field of education? Such questions are important in contextualizing the project. Also important from an evaluative standpoint are the local relationships between CHILD and its community of stakeholders. What do its consumers—the school districts, the parents, and other referring professionals—believe is CHILD's role in addressing student mental health? What results are they interested in? Where do these external pressures meet with CHILD's own identity and mission? The question, "How does CHILD claim its expertise?" is a bi-directional function of this interaction because they are only as "expert" as their consumers and stakeholders believe them to be. To help organize the various factors and dynamics influencing the methodology of the evaluation, an ecological model as developed by Bronfenbrenner (1994) and described above will be utilized. First, however, an overview of the project and its activities prior to this point is necessary.

Preliminary Phase

Program evaluation can be conducted internally or externally, and according to Fitzpatrick et al. (2011), there are potential strengths and limitations to both. This project began during the evaluator's therapy practicum at CHILD, effectively establishing it as an internal evaluation in its early phases. By year's end, however, the evaluator's practicum had concluded, thereby shifting the role from internal to external evaluator. This shift, and capacity as both an internal and external evaluator, has had some implications for the project. As pointed out by Fitzpatrick et al., evaluators with internal experience tend to have greater knowledge of the program and its model, its personnel, and its history. They experience its daily operations in an ongoing fashion, and as a result have a more intuitive understanding of its value system and the impact this has on the students. This familiarity or intimacy with the program can foster a heightened sensitivity to the needs of the program, and predicts greater engagement with and loyalty to the stakeholders.

The downside to this arrangement, however, is the potential for internal evaluator bias. According to Fitzpatrick et al. (2011), there is a tendency for internal evaluators to produce overly positive reports in collaborations with the Executive or Leadership Team. Objectivity is compromised by virtue of the fact that the internal evaluator is in essentially a dual role with the organization—the evaluator has a stake in the program's performance and so the likelihood of a clear and accurate depiction of the program under consideration may be reduced. In the case of the present evaluation, this became less of a concern as the evaluator shifted into an external role with the school. A greater degree of distance emerged between the evaluator and the program, allowing for greater

objectivity. At the same time, the initial internal role provided an intimate glimpse at the interior functioning of the school, enhancing the evaluator's perspective and understanding of CHILD and approach to this evaluation. Thus, the project in some ways afforded the benefits of both an internal and external role, while minimizing the downsides of both.

After initiating the project with CHILD's Leadership Team in the fall of 2012, a series of conversations led to a basic direction for the project. The initial evaluation question arose, "How does CHILD claim its expertise?" From here, it became clear that the Leadership Team was interested in understanding better, both for its own sake and for its consumers, the impact CHILD has on student outcomes. Information or data of this nature could then be used to help establish a claim, or claims to expertise. As their programming is designed to address a number of areas, a decision had to be made as to the scope and direction of the project. After an initial dissertation meeting in June of 2013, it was decided that more information should be gathered from other CHILD stakeholder groups. With approval from the AUS institutional review board, an informal survey (see Appendix B) was created to distribute to other groups affiliated with CHILD, including the teachers and staff, the board of trustees, referring professionals, and parents of former students. At the time the survey was distributed, a number of avenues of exploration were still under consideration, including an evaluation of model fidelity. The survey's main purpose though was to briefly assess for the perceptions of other stakeholders. While not statistically reliable (about 20 surveys out of 60 were returned, but different stakeholder groups were not evenly represented, and the circumstances around data collection varied considerably between groups), two themes emerged in the

survey responses. First, most respondents were exceedingly positive in their attitude towards CHILD. Nearly every survey expressed some level of gratitude and/or appreciation for CHILD and their services. Secondly, most respondents had a fairly clear idea of what CHILD is attempting to do, and how they are attempting to do it. This was especially true for the clinical and prevention teams, who work directly with the students. These groups also had a near 100% response rate on the survey, compared to less than 10% amongst teachers, instructional assistants, and referring professionals. Those responsible for service delivery were similar in program knowledge and attitude, suggesting a high level of model fidelity.

At the same time, ongoing meetings with members of the Leadership Team revealed an interest in student mental health outcomes, in particular. While the survey results above demonstrated that program staff were functioning more or less on the same page, indicating good model fidelity, there was no information indicating the work they were doing was having any of the intended effects. In tandem with this concern was a theme that emerged about reputation in the community. Many of the respondents in this initial survey were concerned that, despite the positive impact CHILD has on its students, much of the broader community was either unaware of the program entirely, or carried misconceptions about the type and quality of the work conducted at CHILD. Outside of their obligatory IEP goal-tracking, CHILD has little in the way of indicators of student progress. The data that CHILD does collect is limited, and at the present moment not integrated into a coherent narrative of the ways their students are served or helped. Given CHILD's mission to "promote social [and] emotional development" amongst its students, and internal concern that external stakeholders and community members might lack

insight into CHILD's impact, it followed that empirical data in this area of mental health would be of benefit to CHILD. In what types of data would external stakeholders be interested? Would this influence their perceptions and opinions of CHILD?

While less salient in the preliminary survey responses, questions and concerns pertaining to long-term program effectiveness of the program also arose. What becomes of CHILD's students? Do they graduate from high school or attend college? Do they find work or secure housing? Do they encounter any legal difficulty? Do they suffer mental health problems? Are they successful in the schools they return or move on to? Aside from the happenstance anecdotal report from a parent or, more rarely, school official, CHILD understands little about the long-term outcomes of its students. The survey of internal stakeholders revealed an interest in this area.

The preliminary survey results from CHILD's internal stakeholders did not raise any significant issues about model fidelity or staff morale, but it did reinforce the Leadership Team's concern with student progress, particularly in the area of mental health. Understanding exactly what sort of impact CHILD has on its students was a prevailing theme. As such, measuring CHILD's impact on student mental health outcomes during their enrollment at CHILD was considered as a possible objective for the present project. However, this was complicated by the fact that CHILD is, on average, a two to three year program from the time of enrollment. As a dissertation coinciding with the evaluator's course of study and training, there were limitations to the length and scope of the study from the beginning. Since an impact study on mental health outcomes would require a period of time beyond what is available for the current project, the objective of the project became a thorough evaluation plan rather than an impact

study. Rather than measuring CHILD's impact on student mental health, in other words, this evaluation seeks to design for CHILD a data-tracking system that CHILD will be able to, of its own accord, implement and use for the purposes of tracking outcomes. No data on student mental health progress will result from the current project, only recommendations for a system to collect such data. Because there is an interest in long-term outcomes as well, questions of this nature were included in surveys for external stakeholders. If these stakeholders likewise expressed an interest in this type of data, it would make sense for CHILD to also track this type of data. Therefore, the development of a data-tracking system across multiple domains is considered, rather than a single area.

The primary audience for the evaluation plan was considered to be the Leadership Team at CHILD, because they are the decision-makers and thus responsible for eventual project implementation. CHILD's board of trustees was also regarded as a potential audience if implementation of the recommended procedures were to require their approval (e.g., to allow additional spending). However, this was not the case. Beyond these two groups, other stakeholders such as teachers and staff, parents and families, school districts and other referring professionals were considered secondary audiences. They are not primary because they aren't responsible for deciding when and how to implement the recommended procedures. However, they may be involved with and/or affected by any program changes and so should be considered an audience of the final report in this capacity.

Data Collection Phase: A Three-Tiered Model

In order to claim expertise in any domain, one must have a sense of the way others define or understand that domain. This is especially the case in the social sciences,

where socially constructed terminology can carry a wide range of meaning depending on the context. As Doll et al. (2012) point out, it may be useful to view mental health in a school context through Bronfenbrenner's (1994) ecological framework. This is the idea that a child's or student's mental health is determined by "multiple tiers of influence" (p. 45): the microsystem, or the immediate environment with which the child interacts, the mesosystem, or the different environmental settings the child moves between (e.g., their school, their home, etc.), and the macro-system, or the broad cultural and political realm that contains within it the other more local layers. For instance, a student's interactions or behaviors at CHILD could be understood as micro-level interactions; their network of family, friends, and other groups and activities could be understood as their meso-level environment; and the influences of law and culture—for instance, the Individuals with Disabilities Education Act (IDEA) and/or national/global perceptions of special education students—constitute their macro-level. As its students' mental health is determined by these multiple levels of influence, so are CHILD's claims to expertise in this regard. In other words, the degree to which they can say with confidence that they are addressing these needs is a function of the way these different "tiers" define and contextualize mental health and their efforts. Therefore, before it can understand its claims to expertise, it must first understand how these systems define expertise. How would members at different levels of that system know, or be convinced, that CHILD was expert in this regard? What types of information are required to understand whether or not CHILD can even claim expertise in the first place? From here, a refined version of the primary evaluation question may read something like, "What kind of data should CHILD be tracking on a regular basis in order to understand the impact it's having on

student mental health?" Put more broadly, it might read something like, "How can CHILD demonstrate effectiveness, and therefore claim expertise in the area of mental health?" If CHILD can demonstrate they are having an impact using data that is not only valid and reliable, but meaningful to interested (and potential) stakeholders, they may be in a position to claim expertise in that area. Because a student's mental health is affected by his or her immediate environment, broader network of social connections and organizations, and even broader cultural environment, information will be gathered from all three levels of CHILD's ecological system: the micro-, the meso-, and the macro.

Using the methods detailed below, data was collected, analyzed, and integrated into a collection of recommendations CHILD could adopt as a part of new data-tracking system. A summary of the survey findings was put into a report and delivered to CHILD's Leadership Team during the evaluation process (Appendix E). Feedback from the Leadership Team at that juncture, and at additional points along the way (per the micro-level of inquiry described below), was incorporated into the final recommendations.

Micro-level evaluation: Dialogue and records review. At the micro-level, three forms of data collection took place: conversations with members of CHILD's Leadership Team, a review of the mental health-related data CHILD currently tracks, and a review of their budget. These were informed dialectically as the project unfolded via the ongoing literature review described in the section on macro-level evaluation. The first of these three was a continuation of what has been an ongoing dialogue since the beginning of the project in the fall of 2012. It was not a formal data collection process per se, but dialogue with the Leadership Team has and continued to be an essential part of the project

throughout. They are the primary decision-makers and program managers, and, as experts of their own programming, were an essential resource when integrating the various aspects of the evaluation. As recipients of the final evaluation report as well, it was critical the results were compiled and interpreted in such a way as to have a meaningful impact on their program and their work. These ongoing conversations were at times recorded with the verbal permission of the participants (unanimous consent was acquired in group meetings), while in other instances notes were taken of the conversation. All recordings and notes were stored on a password-protected hard drive and destroyed at the conclusion of the project.

The second component to the mirco-level evaluation was a records review of the data CHILD currently tracks on mental health-related criteria. This included inquiry into CHILD's process for creating and tracking IEP goals. Other data points such as standardized test scores per annual yearly progress (AYP) requirements were considered as data points to demonstrate CHILD's impact, particularly if schools and school districts are under pressure to ensure AYP of its students. Other data, such as individual student data transferred to CHILD from other schools and programs was considered during the development of a data-tracking system. However, as the primary purpose of the proposed system is to track mental health outcomes, emphasis was on collecting data pertaining specifically to mental health (as opposed to more indirect measures, such as student AYP progress). That being said, the possibility of tracking long-term outcome data on its students was also explored, for the reasons listed above.

Finally, inquiries were made to CHILD's Leadership Team regarding the available budget for implementing final recommendations. Efforts were made in

designing the system to reduce costs wherever possible, given the results of the meso-
and macro-level inquiries described below.

        Meso-level evaluation: Survey distribution. CHILD's meso-level is comprised

of the communities in which they are embedded, not least of which are the consumers

that utilize their services. This base consists primarily of the school districts and families

that enroll children and youth into CHILD. As important stakeholders at this level, it was

important to get input from them regarding student mental health outcomes. CHILD is

committed to the promotion of social and emotional development in its students, but there

is little data in terms of the actual impact their programming has on related outcomes,

both short- and long-term. For instance, while CHILD seeks to address its students'

"lagging skills" through the CPS model, there is no system set up to determine

improvement in this regard. The aim of the project was to understand precisely which

variables along these lines should be measured. To do so, it was important to understand

the perspective of CHILD's consumers. If the outcomes they are interested in are

different from those CHILD is striving for, the potential for client dissatisfaction

increases. Therefore, two surveys were developed to explore consumer perception of and

expectations for CHILD's existing services, including the impact these services have on

student mental health. Created specifically for this project, they were originally designed

to be administered in a cross-sectional design format to CHILD's two primary consumer

stakeholder groups, school district personnel and the parents of students currently

enrolled. However, due to developments in the evaluation detailed below, the two

versions of the survey eventually become two separate, but related surveys. Given the

importance of stakeholder opinion to CHILD's claims to expertise, aspects of these

surveys were adapted for use in the final proposed data-tracking system, and are described in the Results and Discussion, and Recommendations sections below. The following paragraphs explain the methodology for survey distribution as it was administered in the present project.

CHILD has established relationships with each of the school districts they service. Each district has a special education director as well as a representative assigned to CHILD while a student of theirs is enrolled. Both the directors and the representatives were the target population for the school district survey. Since this number is limited to the number of school districts CHILD works or has worked with—roughly 20 and not more than 25—sampling wasn't necessary; all special education directors and school district representatives with whom CHILD has had contact with at some point in the last three years received the survey. Each family with a student currently enrolled received a survey as well. The content of the two surveys used for school district representatives and parents are located in Appendices C and D, respectively.

CHILD provided the names and contact information of both the school district representatives and the families. Given the sensitive nature of their relationships with school district representatives and interests in maintaining positive ties with parents, CHILD's Leadership Team was asked to review the content and questions of the survey before distribution. While the surveys remained largely intact, one major change arose out of this review. It became clear at this point that CHILD was interested not only in claiming expertise in the area of mental health, but more directly in understanding the opinions of external stakeholders, particularly school district representatives. Therefore, a number of changes were made to the school district version reflecting this interest.

Instead of trying to determine how perceptions changed across time, as in the parent

version, the final version of the school district survey was oriented more towards the

opinions of the representatives, the likelihood new outcome data would impact their

decision to send students to CHILD, and their preference for CHILD in relation to other

schools of its type in the area. This led to substantial alterations to their version of the

survey, diminishing the ability of the evaluator to make direct comparisons between the

two stakeholder groups.

The number of potential survey respondents was small, and so it was important to

develop and distribute the survey in a way that minimized potential for nonresponse error

and maximized overall response rates. This approach was informed by a social exchange

model, which states that, "people are motivated to act by the benefits they expect to

receive" (Dillman, Smyth, & Christian, 2009, p. 23). Thus, the likelihood a given survey

respondent would complete a survey was understood as a function of a cost/benefit

analyses: what's in it for the respondent? Social exchange depends on the establishment

of trust, or the basic belief that one's action towards another (cost) will result in a fairly

reciprocated behavior (reward). By sending someone a survey, one is in essence asking

that person to invest certain amounts of time and energy. Thus, a low response rate is

predicted if the respondent doesn't see how expending these resources might result in

reward (Dillman et al., 2009). To account for this, a number of strategies were employed.

First of all, it was important to explain to the respondents the nature of the survey

and the potential benefits of its completion (both to CHILD and to the respondent), and to

do so in a way that subordinated the survey sponsor to the respondent. This can appeal to

their sense of social commitment without putting them in a position of dependency on the

sponsor. Secondly, gestures of appreciation and positive regard act as social reward, and can improve the respondent's sense that their responses are important and worthwhile. This can be heightened by demonstrating an appreciation for the respondent's value system, or by likening survey response to an act of community investment. Both of these principles were incorporated into the tone of a letter sent out to all respondents two weeks prior to the beginning of survey administration. In addition, respondents were also given a five-dollar gift card for use at a local coffee vendor upon completion of the survey, a fact noted in this same letter. Finally, an effort was made to reduce perceived cost in the exchange by minimizing requests for personal or sensitive information, by making it convenient to respond, and by keeping surveys short and simple to complete. These tenets were incorporated into the survey composition administration itself.

The surveys themselves were comprised of quantitative and qualitative data, open- and closed-ended questions. Due to logistical and practical considerations between the two distinct populations, two different surveys were created. For instance, Question 1 on the school district version asks how long the respondent has known about CHILD, whereas Question 1 on the parent version asks what their relationship is to the child that is enrolled (e.g., Mother, Father, Legal Guardian, or Other). Also, where parents are asked to rate their child specifically on some scale, school district representatives are asked to consider an average, or aggregated version of the children under their supervision that have been enrolled at CHILD. Finally, where the content of the question is more similar, wording has been changed to address the discrepancy in roles between the two groups. For instance, "When your child first enrolled . . . ," is not a phrase

appropriate for school district representatives. Overall, every attempt was made to retain similar structure and items in both surveys for ease of comparison.

There were in general three loosely defined sections to the surveys. The first part, questions 1–6, asked about demographics and respondents' general connection to CHILD. The middle part, (questions 7–15 in the school district survey and 7–17 in the parent survey) asked directly about the respondents' experience with CHILD, including their expectations and perceptions in the area of mental health. Finally, the last section of the surveys (questions 16–28 in the representative survey and questions 18–27 in the parent survey) pertained generally to future functioning of CHILD's mental health programming. The primary purpose behind these questions was to understand consumer perceptions and opinions of CHILD, whether improved data collection would impact these perceptions and opinions in the future, and what types of data, specifically, would be most likely to do so.

The final versions of the surveys, which were subject to a secondary IRB approval process, reflected the changes suggested by CHILD's Leadership Team described above. It also incorporated other minor changes suggested by a panel of independent reviewers, considered to be subject-matter experts in the field. Once the content was IRB-approved, it was arranged into its final distributable form on Survey Monkey, a free, online survey service. As a web-based survey, it was distributed via e-mail with the URL link embedded. Follow-up notices were sent, although depending on the needs of the situation, discretion was used in determining their frequency and content. As Dillman et al. (2009) point out, it's important to vary the messages across contacts with respondents. This decreases nonresponse error because it demonstrates greater interest and

engagement with the process. This was observed throughout distribution of the survey and follow-up contacts with survey recipients to ensure prompt completion of the survey. After the initial mailing, recipients were called, and then provided with additional electronic reminders via the Survey Monkey service.

Macrolevel evaluation: Ongoing literature review. A major part of the evaluation involved the integration of the literature in special education, mental health, and program evaluation with the other aspects of the project. Crucial to development of an effective data tracking strategy or program is a thorough understanding of the relevant macro-level factors and their implications for program success. To this end, additional literature was sought out as necessary, most notably in the search for an adequate mental health measure. As demonstrated below in the Results section, which illustrates and explains the results of the meso-level survey, external stakeholders expressed a strong interest in objective outcome data regarding student mental health. Therefore, a literature review was conducted to determine the best possible measure for use at CHILD as a part of the proposed data-tracking system. This developed as sort of a secondary data collection phase, and is entitled Phase II in the Results section. Other literature was reviewed as necessary to inform any other relevant aspects of the data-tracking system.

Results and Discussion

The Results and Discussion section have been combined due to the unfolding nature of the program evaluation. Each level of inquiry into the ecology of student mental health at CHILD, from micro-level dialogue with CHILD's Leadership Team, to the meso-level survey administration and macro-level literature review connects with and has an impact on the other tiers of inquiry. For example, the results of the meso-level survey administration, presented first below, directly influenced the direction and scope of the macro-level literature review, which is presented second. The micro-level inquiry, which in addition to conversations with CHILD included an investigation into their current data-tracking practices and potential budgetary constraints, is presented last in this section because the final shape and dimensions of the recommended data-tracking system depended on CHILD's response to the results of the survey and literature review. This order of presentation of the results also coincides generally with the chronology of events as they unfolded, from survey administration, to literature review, to micro-level inquiry and dialogue with CHILD. The only major deviation from this sequence was the presentation of survey results to CHILD's Leadership Team immediately following the administration and conclusion of the survey.

The section following the Results and Discussion section, entitled Recommendations, will summarize the findings presented below and detail the specific recommendations for the data-tracking system. The last section, the Conclusion, will summarize the project from start to finish, contextualize the findings, and provide suggestions for future research and evaluation.

Meso-Level Evaluation: A Survey of CHILD's External Stakeholders

To begin to answer the question, "How can CHILD demonstrate effectiveness, and therefore claim expertise in the area of mental health?" it is important to understand the perspective of those utilizing their services. CHILD may already believe they have expertise in this area, but if this opinion is not shared by the consumers of their services, in this case the parents of its students and representatives from their contracted school districts, then any claim to expertise will ring hollow. Therefore it was determined necessary to understand more clearly the perspectives and opinions of these stakeholders. What is their current relationship with CHILD and did this relationship come to be? What are their current perceptions of CHILD's functioning, in general and in the area of mental health? Would data providing evidence of progress in the area of mental health influence these perceptions and opinions? If so, what types of data would they be most interested in? All of these questions were put, in various forms across two different surveys, to the parents of students currently enrolled at CHILD, and the representatives of the home districts of these students.

Survey results: Parents. Questions on the parents' survey were designed with several objectives in mind. Broadly speaking, the aim was to understand more clearly their perceptions of CHILD, their level of investment and interest in their children's mental health, and the degree to which they would be interested in or willing to participate in future data collection programs in this area. The survey opened with demographic questions asking about the parents' history with CHILD (e.g., when they first heard about CHILD, how long their child has been enrolled, etc.) and then transitioned into questions about parent perspectives on their child's mental health. The

middle portion of the survey was concerned with parental expectations for their child's

progress while enrolled at CHILD, and their perceptions or beliefs about the school's

intent and ability to address the mental health needs of their child. The concluding

portion of the survey was oriented to parents' interests in data demonstrating CHILD's

impact on student mental health, the types of data they'd be willing to share as a part of

related programs in the future, and their overall level of satisfaction with CHILD's

services.

Response rates for the parent survey are located in Table 1. These data

demonstrate a strong turnout by survey respondents, most notably those numbers found

in the column on the right. Nearly three fourths of the students had at least one parent

participate fully in the survey, indicating positive stakeholder buy-in to the survey and

substantiating the findings detailed below.

Table 1

Parent Survey Response Rates

	Number of Parents	Number of Households	# of Students with at Least One Parent Represented
Surveys Administered	71	48	46
Number of Responses	41[1]	35	34
Response Rate	58%	73%	74%

Parents were then asked to rank the perceived needs of their child upon initial

enrollment at CHILD in five areas: learning/academic achievement, sensory processing,

[1] Four additional surveys were partially completed; their responses were included in the final data set. Respondents on both surveys were also given the option to skip some of the survey items, which explains the varying number of responses between them.

mental health, problems of daily living such as getting dressed, eating properly, and personal hygiene practices, and speech and language. Using a Likert scale with one (1) being the highest score or greatest need, and five (5) the lowest rating or need, the average rating of each of the five areas are listed in Figure 1. With an average ranking of 2.21, it is clear that parents believed their child's mental health was the area of greatest need upon enrollment at CHILD. Not only does this reinforce CHILD's desire to develop expertise in this domain, but it supports their stated mission of addressing "social and emotional" health in addition to academic-related concerns.

Parent Perception of Child's Needs upon Enrollment

Area	Rating
Speech and language	3.68
Problems of daily living (e.g. getting dressed, eating appropriately, personal	3.68
Mental health	2.21
Sensory processing	2.35
Learning/academic achievement	2.58
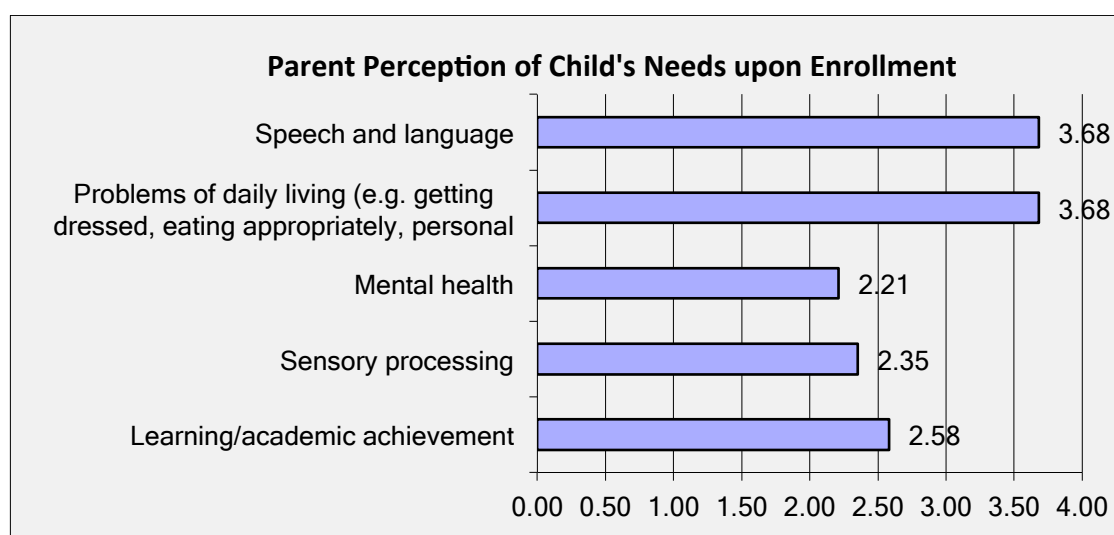
Figure 1. Parent perception of child's needs upon enrollment.

Parents were also asked to rate the severity of their child's mental health needs specifically upon enrollment at CHILD, the results of which are presented in Figure 2. These results indicate that 86% of responding parents believed their child had "moderate" to "very high" mental health needs upon enrollment at CHILD. Perhaps because mental health was identified as an area of such high need, parents also reported high investment in the mental health of their child, as depicted in Figure 3. It is clear from these results

that, at least in the eyes of the parents, mental health is an area of high need amongst CHILD's student population, and of high importance to the students' parents.

In an attempt to understand external perceptions of CHILD's programming, parents were asked to describe their expectations for their child's progress, both in general and in the area of mental health specifically when they first enrolled at CHILD. This was done in part to evaluate how parent perceptions change across time, once they become more familiar with CHILD over the course of their child's enrollment. Data of this nature, while not directly pertaining to student mental health, is intended to help CHILD understand what its external stakeholders perceive to be the areas of expertise in their programming. The questions here were two of several open-ended questions on the
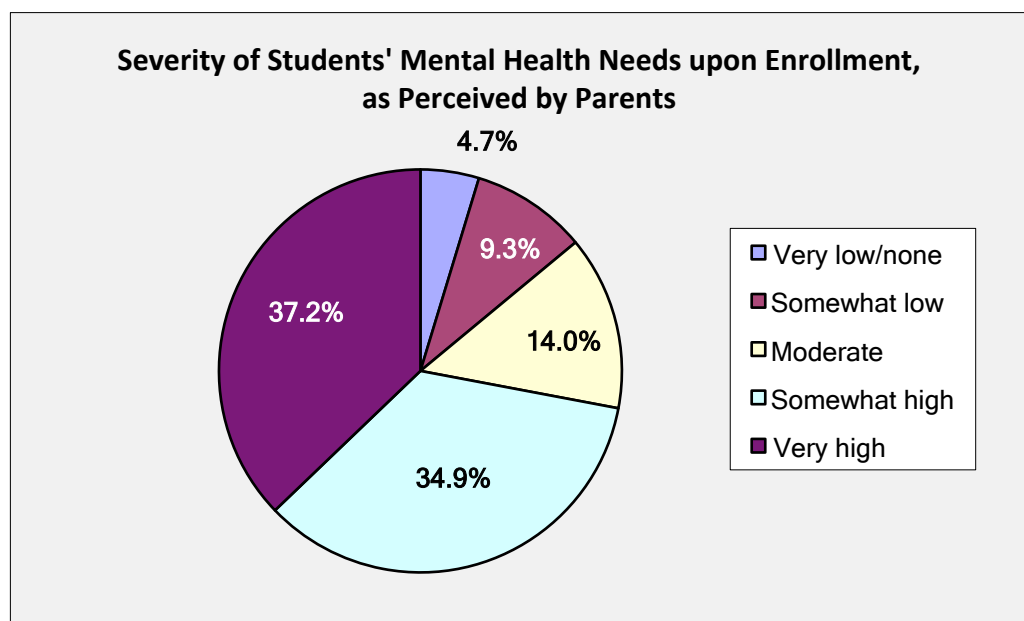
Figure 2. Severity of students' mental health needs upon enrollment, as perceived by parents.

survey, necessitating a process of qualitative analysis for the purposes of extracting common themes in the responses. In regards to general expectations, the most salient theme to emerge out of the 40 responses was that of "social development." Responses in

this domain were typified by expectations such as "learn[ing] to cope better around other people outside of the family." These answers were differentiated from other responses

Parental Investment in Their Child's Mental Health

2.2% 0.0% 2.2%

22.2%

73.3%

Legend:
- Not invested
- Slightly invested
- Moderately invested
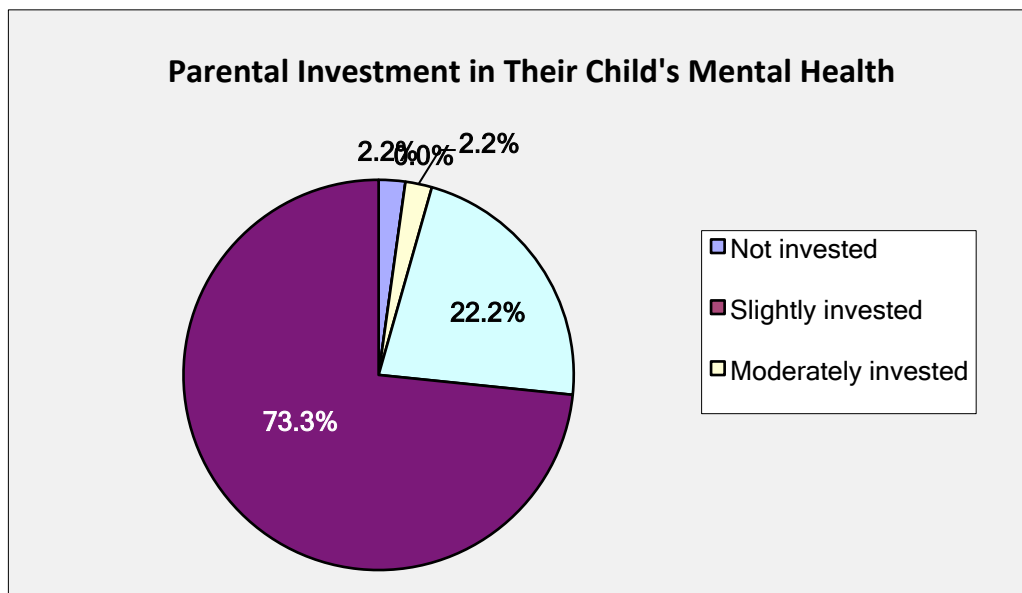
Figure 3. Parental investment in their child's mental health.

oriented towards different but related issues such as mental health and behavior by an overt reference to aspects of their child's functioning in the context of other people (e.g., feeling isolated, interpersonal difficulties, etc.). Conversely, mental health and behavior were treated as a distinct but unified category, in part because these two were often used interchangeably in the responses, and also because these responses did not make reference to a child's functioning in a social context. This theme of "improved behavior/mental health" was the second most common, and was typified by responses such as, "wanted [the child's] mental health to remain stable . . . but also wanted [the child] challenged." The other themes to emerge, though less salient, included "academic concerns," "keeping the child safe," and "[the child] being accepted." This last theme has clear implications for a child's health, and could perhaps be lumped into the category of "social development." However, the point here is to note the overall presence of mental

health as a theme in the parents' responses, relative to school- or academic-related concerns. The question was asking simply for the parents' general expectations for their child's progress while at CHILD, and the vast majority made at least some reference to aspects of their child's mental health.

Responses to the question about expectations for progress in the area of mental health were much more varied. However, two themes tended to appear more frequently amongst the 40 responses: "self-regulation" and "social skills." Those responses labeled as "self-regulation" made reference to mood, frustration tolerance, or emotional regulation, and were typified by statements such as expecting the child to "learn how to manage her anxiety," to "self regulate emotions [and] control his anger [*sic*]," "to learn emotional regulation skills," and to have "self-regulation of behavior and mood [and] reduce aggression." Responses in the "social skills" category made reference to a child's functioning in the context of relationships, such as in communication or feeling accepted by others. Parental expectations sorted into this category included statements such as "learn[ing] to cope better around other people outside of his family," "relate to his siblings and parents lovingly," and "social skills is the main thing he needed." Other themes appeared throughout the responses, although on a less prevalent scale. These ranged from behavior (in general), to overcoming or recovering from trauma, to building trust and a sense of self-esteem or well-being, and to having an opportunity to grow in a stable environment.

The themes described above are not intended to provide a comprehensive and precise analysis of parental expectations for student progress when enrolling at CHILD. They are, however, useful in understanding more generally the perspectives of the parents

as they enter into a relationship with CHILD, and some of the expectations they may have for their child's progress while enrolled there. As CHILD considers implementing a data-tracking system, data of this type will help CHILD understand better the mentality and outlook of these external stakeholders, so that the most appropriate data are collected and presented in a relevant way. This is not to say that the types of data collected and presented should be directly or perfectly tailored to the expectations of the parents (or other external stakeholders) per se. However, gaining perspectives of this sort can help CHILD work to manage stakeholder expectations in a way that minimizes negative opinion or reaction to their service outcomes and bolsters their claims to expertise, especially in the area of mental health.

The next portion of the survey asked parents about their perceptions of CHILD's intent and ability to address the mental health needs of the students, and how these perceptions have changed across time as parents became more familiar with CHILD and their programming. Almost all parents (37 out of 40, or 92.5%) reported their understanding of CHILD's intent to address student mental health ranged from "moderate" to "very high" upon initial enrollment. This rate remained generally consistent when parents were asked to reconsider this question based on what they've learned about CHILD across time. Upon initial enrollment, the parents' understanding of CHILD's ability to address mental health was similarly high: 38 out of 40, or 95% rated it at least "moderate," with 62.5% rating it "high" to "somewhat" to "very high." However, parents reported their understanding of CHILD's ability to address mental health did increase as they learned more about the school and its programs across time. Parents were also asked to rate their child's level of mental health needs, both upon

enrollment at CHILD and currently. Their responses are presented and compared in

Figures 4 and 5 respectively. So while no parents rated their child as having little to no

mental health needs at the time of the survey, in general the perception shared amongst

parents is that their children had fewer mental health needs at the time of the survey then

they did upon enrollment at CHILD.
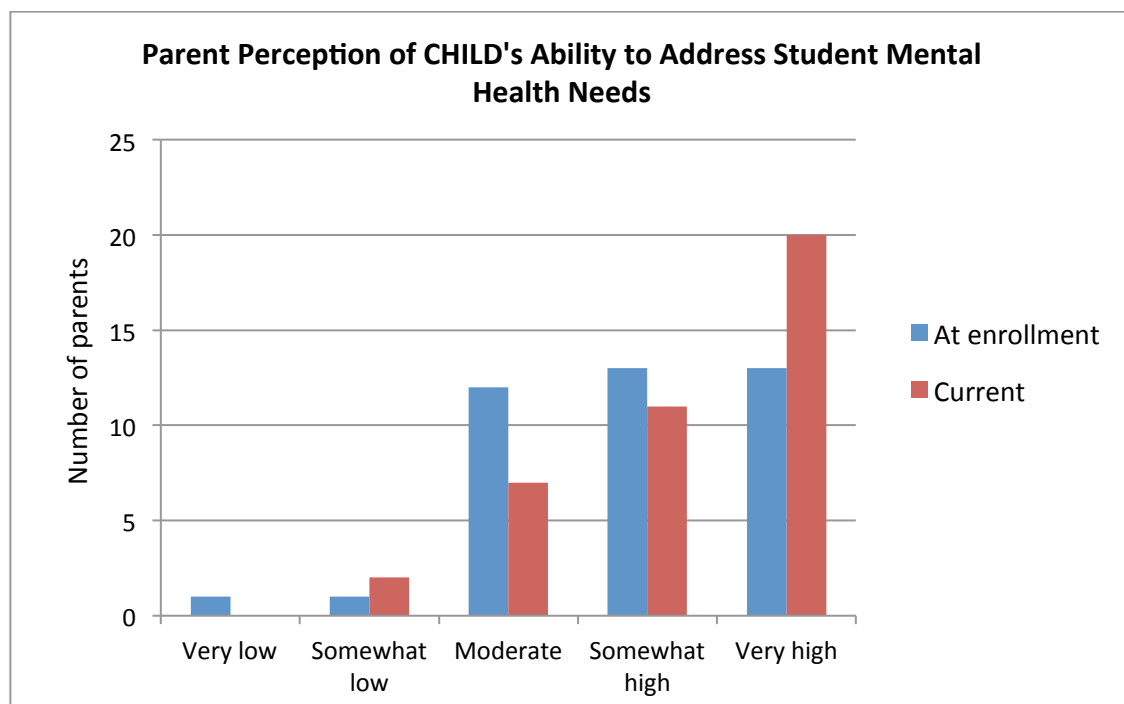
Figure 4. Parent perception of CHILD's ability to address student mental health needs.

These results suggest that (a) parental perceptions of CHILD's intent and ability

to address mental health needs may improve slightly across time, (b) most parents regard

CHILD's investment and skill in working with student mental health positively, and (c)

students' mental health needs, in the eyes of their parents, tend to decrease over time

while at CHILD. While this is not objective evidence of improved student mental health

outcomes, it suggests CHILD is having some positive impact, at least in the minds of the

parents. It also confirms CHILD's assumption, based on anecdotal feedback over the
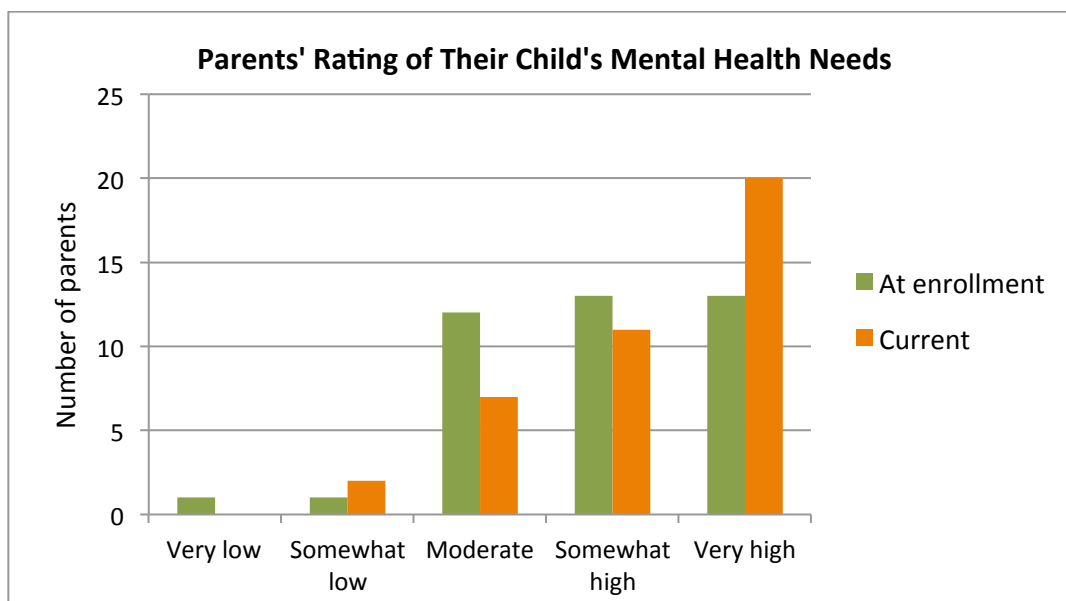
Figure 5. Parents' rating of their child's mental health needs.

years that parents are generally satisfied with CHILD, and tend to hold their services in high regard. As CHILD is concerned with its claims to expertise, data of this nature may be useful to CHILD going forward as it lends some additional empirical support to this long-held belief.

As far as data regarding student mental health at CHILD, parents were asked about areas of functioning they would be most interested in hearing about in the future. Respondents were asked to choose from seven listed domains, with an "Other" option for adding additional domains of interest. Respondents could check as many as they liked. The domains provided were designed to be taken at face value, with no further parameters or definitions provided to the respondents. As such, the data generated by this question should be regarded as general approximations of stakeholder interest, rather than exact representations. Results are presented in Figure 6. The three types of mental health data of greatest interest to parents, based on these responses, are self-regulation, behavior, and coping skills.

The survey also indicated to parents that CHILD is interested in learning more about the long-term outcomes of its student population, such as high school graduation rates, employment and housing status, and college enrollment and graduation. Parents were asked about the likelihood they would be willing to participate in a program designed to track these outcomes. Results are presented in Figure 7. None of the parents said they would be either moderately or very unlikely to participate, and 90.2% said they would at least be moderately likely. Furthermore, when asked which types of data
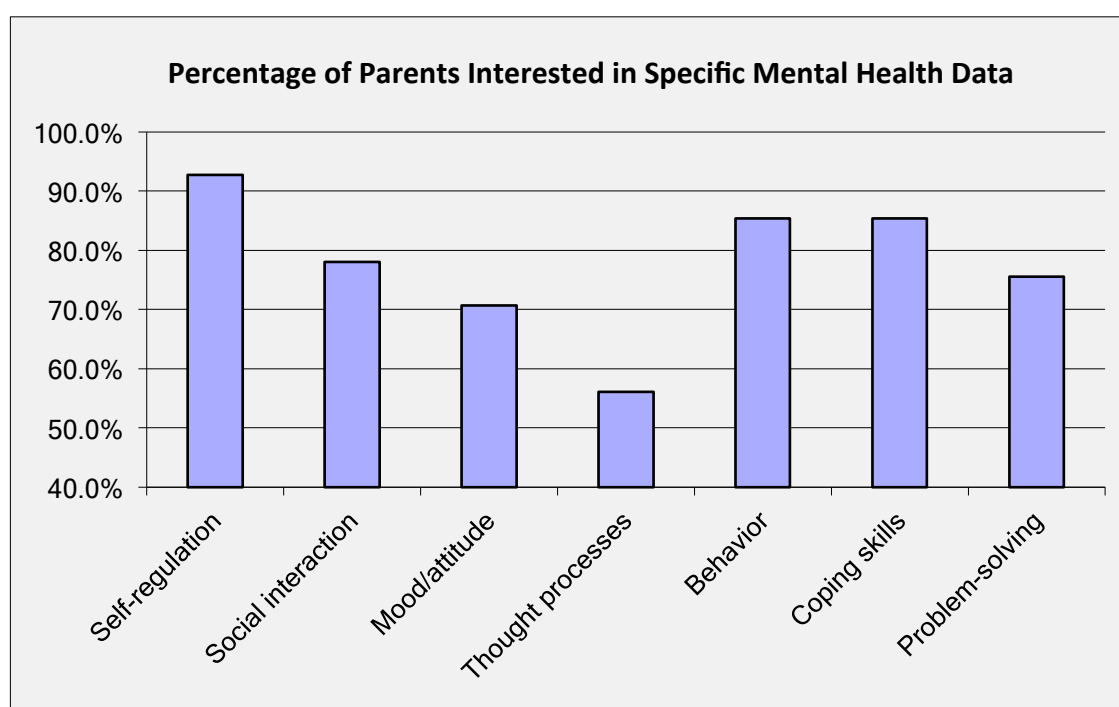
Figure 6. Percentage of parents interested in specific mental health data.

specifically they would be willing to share, most respondents reported everything from high school graduation/GED status to housing status. Several even indicated they would "share everything" or "do whatever [they] can to help CHILD." While most of these data points do not pertain directly to mental health, long-term student outcomes can serve as indicators of general student progress after they have transferred out of or graduated from

CHILD. This information could be used by CHILD, not only to help support their claims to expertise in general, but also as indicators of general student well-being after enrollment in CHILD.
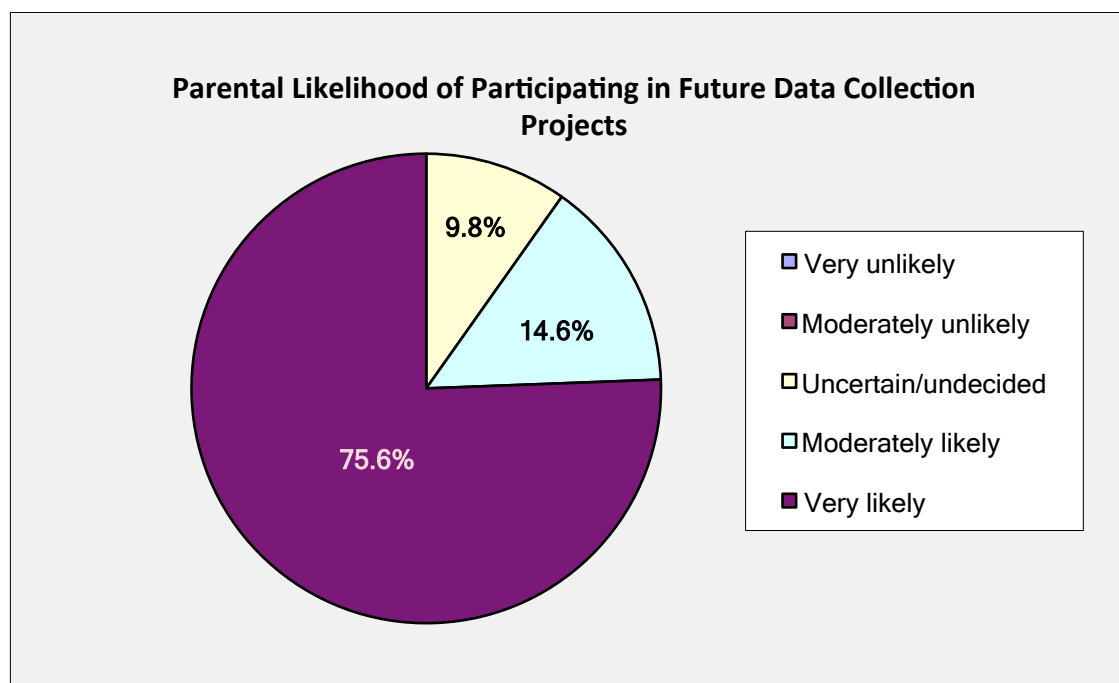
Figure 7. Parental likelihood of participating in future data collection projects.

In the last section of the survey, parents were asked to share their opinions of CHILD across three different questions: two pertaining directly to levels of satisfaction, and a third inquiring about CHILD's contribution to their child. The first two questions asked about overall parent satisfaction levels and then satisfaction levels related to their mental health services. Because these same two questions were asked on the school district representative survey, responses between the two stakeholder groups to both questions have been reflected and compared in Figures 8 and 9, respectively. These graphs indicate that parents hold CHILD in high regard, more so than district representatives. One interpretation of this data is that as external stakeholders become more familiar with CHILD, their opinions tend to improve. Also notable is parent

satisfaction with CHILD's mental health services, which is slightly lower than their

overall satisfaction. While it's unclear exactly why this may be, some responses in the

open-ended questions indicated that a small portion of parents may not believe CHILD's

mental health services are of benefit to their child because their child has low to no

mental health needs. Additional support for this interpretation is evident when comparing

the percentage of parents with lower levels satisfaction with the percentage ranking their

child as having lower mental health needs. As noted above, 28.0% (12 out of 43) of

Figure 8. Overall level of stakeholder satisfaction.

parents described the mental health needs of their child as ranging from "moderate"

(14.0%) to "somewhat" (9.3%) and "very low" (4.7%). This mirrors closely the

percentage of parents whose satisfaction with CHILD's mental health services ranged

from "moderate" (19.5%) to "somewhat" (4.9%) and "very low" (2.4%). Regardless,

parent satisfaction with CHILD and their mental health services appears to be high, and

confirms for CHILD what has long been believed to be the case.
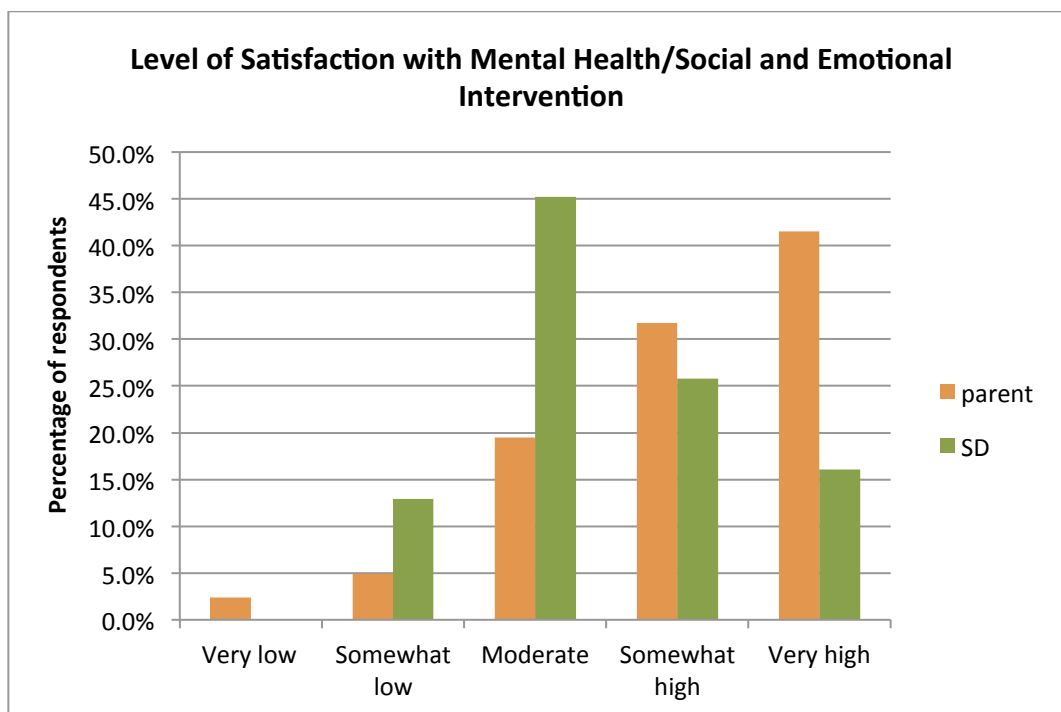
Figure 9. Level of satisfaction with mental health/social and emotional intervention.

The third question pertaining directly to parent opinion was also the final question on this survey. It was designed as an open-ended question asking parents what they believed to be the best contribution CHILD had made to the life of their child. 100% (41 out of 41) offered a response, all of which were analyzed and grouped according to emergent themes. The most dominant theme, appearing in 49% (20 out of 41) responses, was the impact CHILD has on its students' self-esteem and sense of acceptance. Responses in this category were typified by the following statements: "he has rediscovered that he has worth"; "he only associated school with being in trouble. Now he absolutely identifies himself as a student and a learner"; "she has learned that she can be successful in the classroom"; and "he is not alone with his struggles." It also became clear that parents believe CHILD's staff members bring a level of understanding,

encouragement, and advocacy to their work that has a lasting impact. This theme appeared in 26.8% (11 out of 41) of the responses, and was typified by statements such as: "Our son feels love and devotion from his teachers," "[the teachers are] able to deal with behavior issues in a positive manner," and "he isn't worried about how staff will treat him on a daily basis."

Survey results: School district representatives. As a therapeutic day school, CHILD exists to serve those students deemed in need of supports beyond what is available to them in their "home" schools. Parents and school district representatives both are members of each student's IEP team, and as such carry decision-making power in the placement process. However, their roles are different. Once a need has been determined, it falls on the student's school district to facilitate and/or arrange placement with available programs or schools. While the final decision is made by the IEP team, the opinions and input provided by the school district representative assigned to the team impact the final outcome considerably. Therefore, they were identified by CHILD as the other external stakeholder group most relevant to the needs of the current study. The primary purpose of the representative survey was to understand more clearly their perceptions and opinions about CHILD and its programming. In order to provide a comparison between this stakeholder group and the parent group, the initial version of the representative survey was designed to parallel as closely as possible the parent survey. The final versions of the administered surveys were similar in content, and shared several of the same questions. However, due to fundamental differences in the nature of the representatives' relationship with CHILD's students, as well as their unique role in the student placement process, many of the questions were either altered or changed entirely.

While both parent and school district surveys were concerned with perceptions and opinions of CHILD, for example, the school district survey did not ask representatives if they would be interested in sharing post-enrollment information with CHILD because they simply wouldn't have access to that information. When between-group comparisons are available and relevant, a chart or table is provided to illustrate the information. Otherwise, the information provided below by district representatives is independent of the parent data, and should be considered in light of their role as partners with CHILD, sometimes for multiple students.

As with the parent survey, the aim with this survey was to understand more clearly district representative perceptions and opinions of CHILD and to gauge their interest in potential future demonstrations of objective mental health outcomes. The first part of the school district survey was concerned with the history and nature of the respondents' relationship with CHILD, similar to the parent survey. Questions in the next phase sought to understand the factors that influence representatives' opinions of programs such as CHILD, and the degree to which CHILD rates on these factors. Given the different, and more distant relationship with the students they are responsible for (relative to the parents), questions about their perceptions of student mental health needs were omitted. The last component of the survey consisted of questions on satisfaction (as with the parent survey), opinions of CHILD's ability to impact student mental health, willingness to send students to CHILD, and ranking of CHILD against other similar programs in the area. There were also several questions exploring interest in objective evidence of student mental health outcomes, and the impact this type of data would have on their perceptions and opinions of CHILD.

Finally, as noted in the Methods section, the phrase "mental health" was replaced throughout the survey with "social and emotional" or "social and emotional health." This was due primarily to CHILD's belief that "mental health" would be interpreted by district representatives in their role as something specific rather than a broad categorization of the topic. In their experience, the use of "mental health" tends to, in the minds of the district representatives, connote and become synonymous with the provision of formal counseling services, which often come at additional expense to the district. "Social and emotional" was considered to be an adequate substitute due to its acceptance on a national scale as a proxy for mental health (e.g., as evidenced by the prominence of organizations such as the Collaborative for Academic, Social, and Emotional Learning, or CASEL), and in its alignment with CHILD's stated mission of addressing student social and emotional needs.

The survey was distributed to a total of 46 school district representatives across 31 school districts. The original list of representatives provided by CHILD required some updating due to staff turnover on the district end. Some districts had multiple representatives, others had only one. Several of the representatives were unavailable for various reasons (e.g., maternity leave) and were not provided a survey; however, all other available and active school district representatives were administered a survey. Of the 46 recipients, 31, or 67.4%, responded in full. Of the 31 school districts, 24, or 77.4%, were represented by at least one respondent. Given the small population of respondents, a high response rate was important, especially in terms of district representation. As with the parent version, which had a higher percentage of households represented than percentage of individual parents, the percentage of districts represented was higher than the

percentage of individual representatives responding. This demonstrates good coverage

across the stakeholder groups, and lends reliability to the results provided.

      Nearly all representatives reported some degree of familiarity with CHILD, as

depicted in Figure 10. Respondents were also asked to review a list of factors they might

consider when determining which program is best for a given student. They were then

<figure>

School Distrcit Representatives' Levels of Familiarity with CHILD

6.5%

41.9%

51.6%

- Not very/not at all familiar
- Somewhat familiar
- Very familiar

</figure>
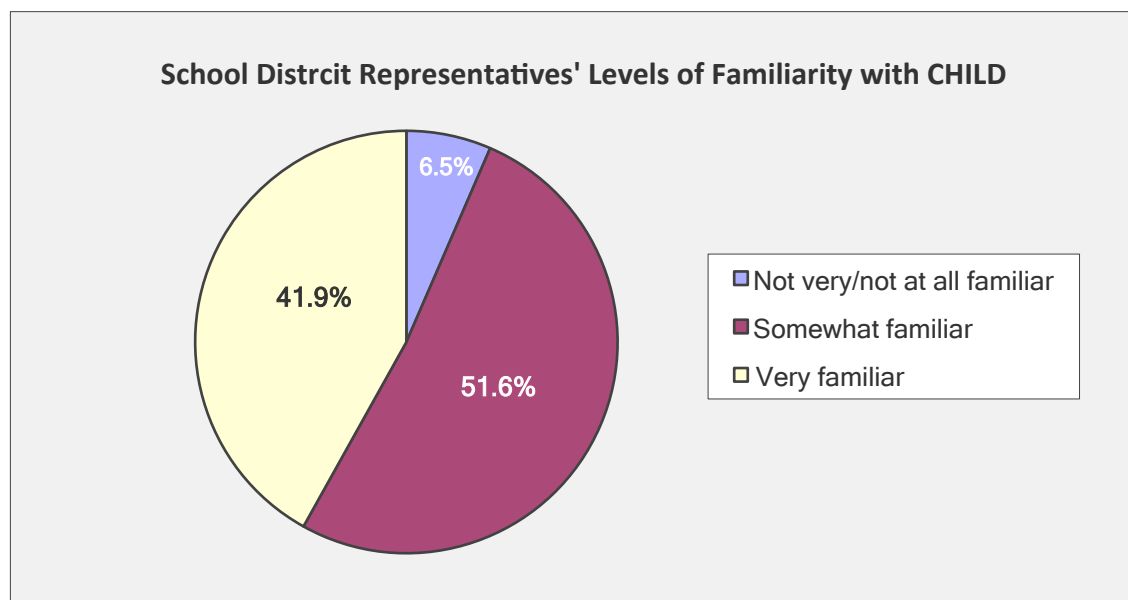
Figure 10. School district representatives' levels of familiarity with CHILD.

instructed to rate each factor on a 4-point Likert scale according to its level of

importance, from "extremely/always important" to "not at all important." The following

two factors were ranked almost unanimously as "extremely/always important": whether

the program is a good fit for the needs of the student (28 out of 31 respondents) and the

program's effectiveness with social and emotional intervention (27 out of 31). The

program's effectiveness with academic intervention was also ranked as

"extremely/always important," but less so than the other two (21 out of 31). Others listed

as important included CHILD's willingness to take new students, and staff expertise. It's

clear from this data that needs beyond just the academic weigh heavily on the placement

decision for the representatives that work with CHILD. Federal legislation under IDEA (2002) makes clear that "an effective educational system serving students with disabilities should . . . address the full range of student needs, particularly the needs of children with disabilities who need significant levels of support to participate and learn in school and the community" (STAT. 2763, C.). Furthermore, it states that a student's annual IEP goals, a core component in their special education services, should be "designed to . . . (aa) meet the child's needs that result from the child's disability to enable the child to be involved in and make progress in the general education curriculum, and (bb) meet each of the child's other educational needs that result from the child's disability" (STAT. 2707, d.1.A.i.II (aa) and (bb)). In other words, school district representatives must consider the needs of a student beyond those specifically academic. Given the high proportion of students with social and emotional needs at CHILD, it appears that representatives are for the most part fulfilling their mandate by weighing CHILD's ability to address these needs. It also underlines the importance of considering the aspects of CHILD's programming dedicated to student social and emotional, or mental health functioning.

Respondents were asked to rate CHILD on the same list of factors as discussed above on a 4-point Likert scale from "not at all appealing" to "very appealing" (with a fifth option, "I don't know/it depends on the student" also available). Based on their ratings, the following were listed as either "appealing" or "very appealing": the fit of CHILD's programming to the needs of the student (22 out of 31 respondents), CHILD's

efforts to reduce/eliminate restraints[2] (20 out of 31), CHILD's effectiveness with social

and emotional intervention (19 out of 31), and CHILD's effectiveness with academic

intervention (14 out of 21). In other words, two of the top three rated factors match the

two factors identified as most important to district representatives in their consideration

of programs such as CHILD: the fit of the program to the student and its effectiveness

with social and emotional intervention. In a separate question, they also ranked CHILD's

effectiveness with social and emotional intervention as the most improved amongst the

factors listed above.

District representatives were asked directly how likely, in general, they were to

send a student to CHILD. The results to this question are presented in Figure 11. It would

appear at face value that a significant portion (over one third) of the representatives are

unlikely to send students to CHILD. As respondents either potentially or actively

engaged with CHILD and their programming, this could be potentially concerning to

CHILD. However, given the percentage of respondents that do not currently have a

student enrolled there (29.0%, or 9 out of 31), this could merely be a reflection of

unfamiliarity with the program rather than a statement on the quality of their program. It

is worth noting too the types of students school districts are likely to send to CHILD

based on special education category (e.g., learning disability, other health impaired,

emotional disturbance, autism, etc.). Respondents were asked to rate how likely they are

to send different categories of students to CHILD, the results of which appear in Figure

12. It is possible that those unlikely to send students to CHILD are unlikely to do so

because it is not a good fit for the students they tend to work with (noting again the

---

[2] The use of restraints and/or seclusion in a school setting has generated considerable controversy over the years. CHILD has made a concerted effort to reduce its prevalence at their school, and is curious about its external stakeholders' awareness of and interest in this effort.
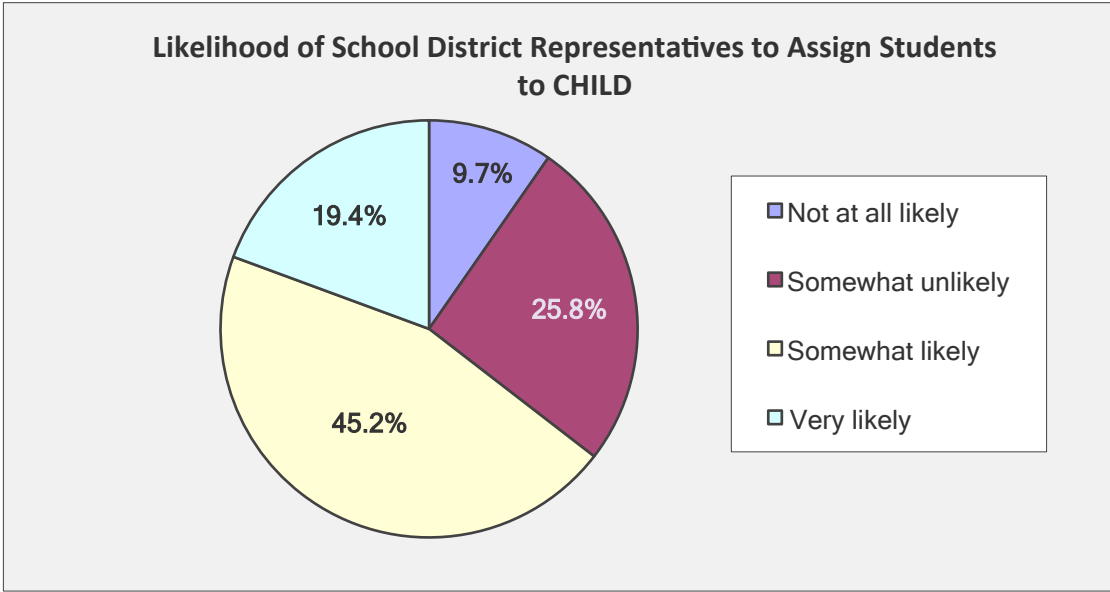
 56

Figure 11. Likelihood of school district representatives to assign students to CHILD.
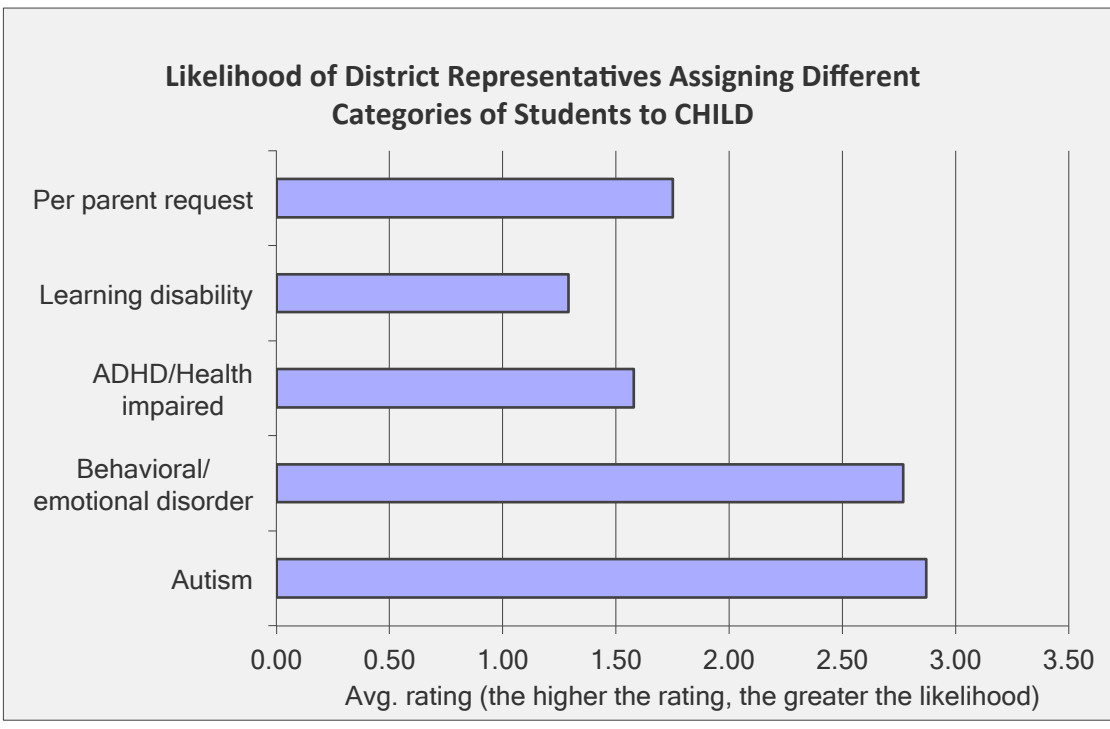
Figure 12. Likelihood of district representatives assigning different categories of students to CHILD.

importance of 'goodness of fit' discussed above). The data also highlight the tendency of district representatives to consider CHILD as a placement for students with an autism spectrum or related disorder.

While it's possible to make inferences from the data presented here, a significant portion of the district representative survey was devoted to understanding directly the opinions and perceptions of the representatives who work with CHILD. Part of the reason for this is the expressed desire from members of the Leadership Team that CHILD be perceived as the top-rated program of its kind in the area. For this reason, one question on the school district version asked respondents to rank CHILD against several other schools in the area, in order of overall preference. Of the 25 responses to this question, it is clear is that one program in the area stands out as more preferred amongst school district representatives (see Table 2).

CHILD was ranked second overall on this list, based on average ranking, but the other two programs ranked a close third and fourth. Of note as well is the difference in modes between each school's response set. The three schools besides CHILD have fairly well-pronounced modes, with 15 of the 25 respondents ranking The Overlake School first, 12 ranking Renton Academy third, and 13 ranking Northwest School of Innovative Learning (NW SOIL) fourth. CHILD's rankings, however, were more evenly distributed across the four response options, with a slight skew to the third and fourth ranks. This suggests a more variable impression of CHILD. Given that favorability may be linked positively to familiarity with CHILD, it's possible that CHILD's ranking could improve with increased familiarity over time.

Table 2

School District Representatives' Rankings of Local Therapeutic Day Schools

Ranking:	1 (highest)	2	3	4	Rating Average
Renton Academy	3	6	12	4	2.68
The Overlake School	15	4	4	2	1.72
Children's Institute for Learning Differences (CHILD)	4	7	8	6	2.64
Northwest School of Innovative Learning (NW SOIL)	3	8	1	13	2.96

Some questions in the survey were included specifically to explore ways CHILD might increase district representative preference for CHILD's services. These included questions about their perceptions about CHILD's present functioning at various levels, particularly around the domain of mental health. For instance, in their marketing materials, CHILD makes the claim that "100%" of their students make "significant emotional, behavioral, and academic improvements." District representatives were asked to rate the degree to which they agreed with this assertion, on a six-point Likert scale. Their responses are reflected in Figure 13. Overall, they tended to agree more than they disagreed, but only by a slim margin. Nearly half said that they disagreed with the statement in some capacity. This is significant because (a) CHILD has no

objective data to back up the claim that "100%" of their students make "significant . . .

improvements," and (b) it suggests that corroborating objective evidence could enhance

stakeholder perceptions in this regard. In the question immediately following,

respondents were asked to rate CHILD's ability to positively impact student social and

emotional health (see Figure 14 for results). Overall, 95.8% of representatives in this

sample believe CHILD has at least a moderate ability to affect positive change in student

social and emotional health. Only a very small percentage (under 10%, not more than

three respondents) rated CHILD as having low ability in the domain of social and

emotional health intervention. As noted above in the parent survey results, district

representatives were also asked to rate their overall levels of satisfaction with CHILD as

well as their levels of satisfaction with the social and emotional (or mental health)

services. Refer to Figures 8 and 9 for the representatives' responses to these questions.
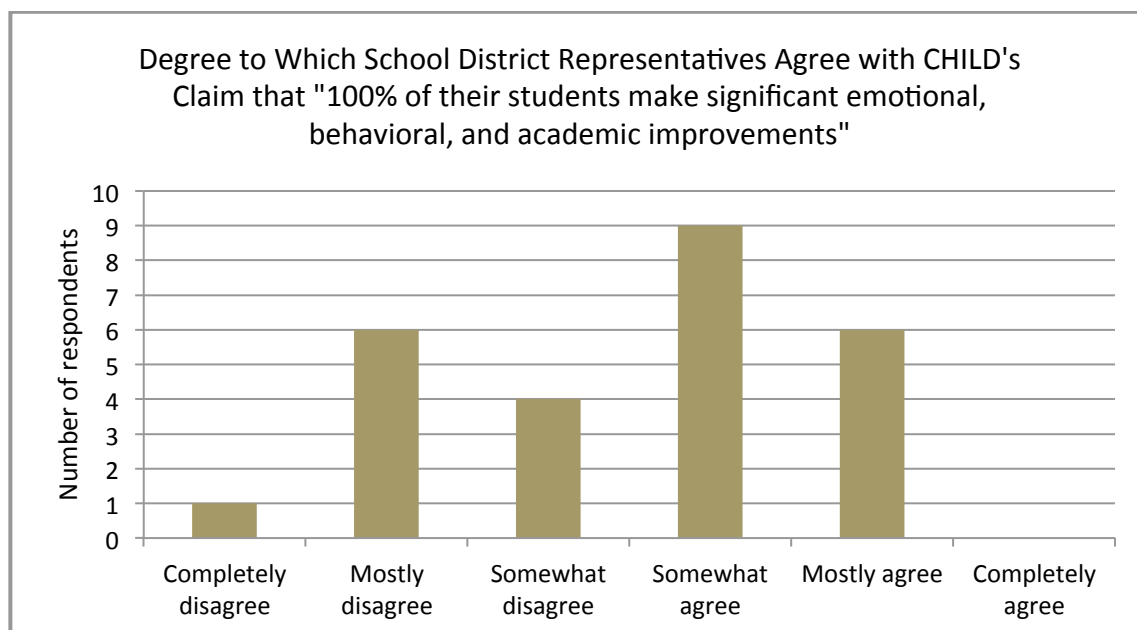
Figure 13. Degree to which school representatives agree with CHILD's claim that "100%

of their students make significant emotional, behavioral, and academic improvements.

School District Representatives' Ratings of CHILD's Ability To Positively Impact its Students' Social and Emotional Health
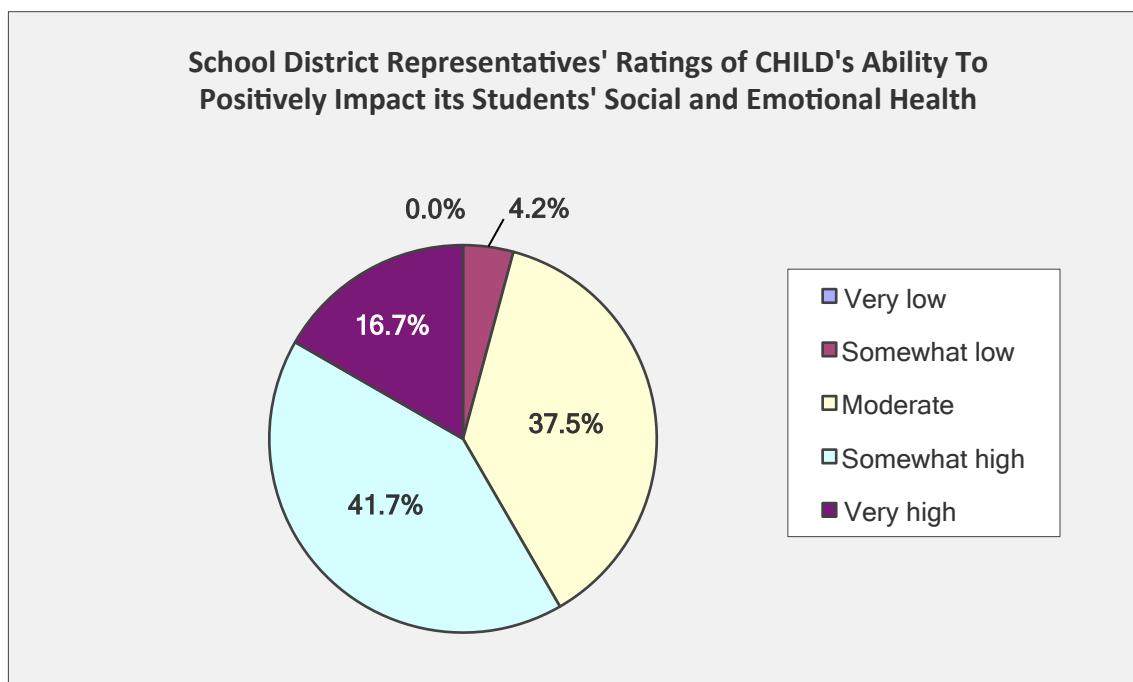
Figure 14. School district representatives' ratings of child's ability to positively impact its

students' social and emotional health**.**

The data indicate that levels of overall satisfaction are nearly identical to their levels of

satisfaction with CHILD's mental health services. They retain a generally positive view

of CHILD, though somewhat more moderate than parents of currently enrolled students.

It is clear based on this data that the school district representatives servicing

CHILD have a positive impression of their services. Yet, there is room for growth,

especially in light of CHILD's desire to rank as the preferred program in the area. To this

end, respondents were asked whether or not objective evidence of student outcomes

would influence both their opinions of CHILD and their willingness to refer students to

them. This part of the survey was broken up into three questions. The first question

asked, "If CHILD was able to demonstrate its effectiveness using quantitative or

empirical data other than IEP data, would this influence your decision to assign them

students from your district?" To this, 71% (22 out of 31) replied "Yes," and 29% (9 out

of 31) replied "No." The wording of this first question was designed to differentiate objective, empirical data that could be potentially collected from that already generated by CHILD in the IEP goal-tracking process. Each IEP goal is reviewed annually and the school or program is obliged to provide evidence of progress (or lack of progress) on each. While CHILD's current system for tracking these outcomes meets the mandates of the IEP process, it is not objectively rigorous due to the fluctuating nature of the system (as discussed earlier).

The second question regarding stakeholder opinion asked, "If CHILD was able to demonstrate the long-term outcomes of its students, such as high school graduation rates or efficacy of intervention in domains such as social and emotional health, to what degree would this impact your general opinion of CHILD?" To this, 64.5% (20 out of 31) said it would impact their opinion "considerably" and the other 35.5% (11 out of 31) said it would impact it "somewhat." Zero percent said it would only impact their opinion a little or not at all. Given CHILD's general interest in stakeholder opinion and specific interest in superior performance, this provides substantial support for the institution of a more rigorous data-tracking system, not only in relation to short-term mental health, but to longer-term student outcomes in general.

The third and final question about possible future data tracking asked, "If CHILD were able to somehow demonstrate the long-term social and emotional outcomes of its students quantitatively, to what degree would this increase the likelihood you would assign students from your district to their school?" The response profile on this question, which was similar to the first two, is illustrated in Figure 15. In other words, 100% of the

representatives responding said their decision to refer students to CHILD would be impacted by long-term, quantitative data about social and emotional outcomes.

The survey concluded with the same open-ended question as the parent survey: "In your opinion, what is the most important contribution CHILD has made to its students?" The wording was altered slightly from "the life of your child" to "its students" to reflect the contrasting relationships between the different groups of respondents and the students. The most common theme to emerge from these responses was a reference to CHILD's staff, their "investment" in student progress, "level of caring," "collaborative" efforts with the school districts, and "respectful attitude" towards and "willingness" to help with challenging students.
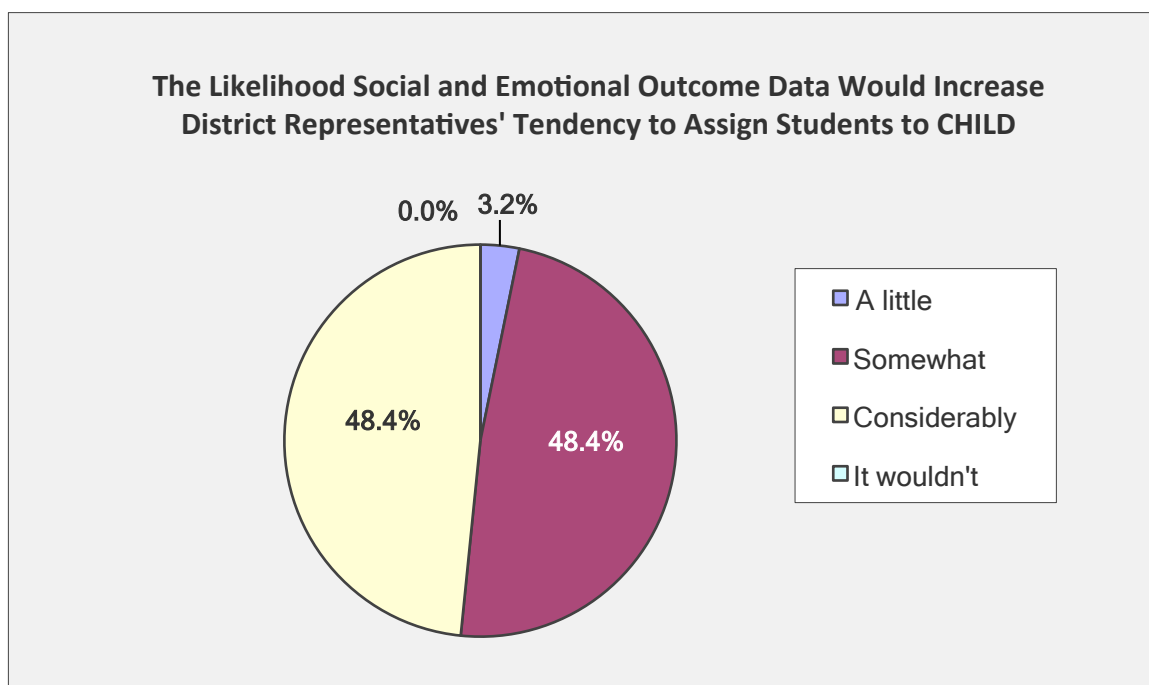
Figure 15. The likelihood social and emotional outcome data would increase district representatives' tendency to assign students to child.

One respondent said that, "at CHILD, the staff are not shocked by the behaviors and understand that it is part of the student's disability and why they are there to help. [*sic*]" Other prominent themes appearing, in order of salience, included CHILD's "willingness to take on tough cases," their impact on student social and emotional health, their ability to tailor programming to the "unique needs of the 'whole' student," their focus on family integration, and their "variety of services."

Summary. To begin to answer the question, "How can CHILD demonstrate effectiveness, and therefore claim expertise in the area of mental health?" it was important to understand the perspective of those utilizing their services. In particular, CHILD identified the parents of its students and representatives from their contracted school districts as the most important stakeholders to this end. The survey results presented here detail responses to questions related to current perceptions and opinions of CHILD's functioning, as well as the impact future data-tracking processes might have on these perceptions and opinions.

In general, the survey findings suggest that perceptions and opinions of CHILD are more positive than negative. Parents and school district representatives both cited CHILD's staff as a prominent strength. Parents, in general, tended to report a high opinion of CHILD, and were most pleased with the impact of their programming on their child's self-esteem and sense of belonging. They would like to see more training for both the school staff and caretakers/families of the students, and would prefer more involvement on the part of the mental health counselors. School district representatives tended to express a good, but more moderate opinion of CHILD, voicing appreciation for CHILD's willingness to take on "tough cases" but a degree of uncertainty over the

effectiveness of CHILD's programming. Responses from district representatives indicated they are more likely to send students to CHILD who qualify for services under the Autism and Emotional Disturbance categories rather than the Other Health Impairment (e.g., ADHD) and Specific Learning Disability Categories, and in general were more likely to rank CHILD behind at least one other program of its type in the area.

In regards to data, parents and school district representatives together expressed a strong desire for more empirical, objective outcome data, especially in regards to students' mental health while enrolled at CHILD. Not only would it likely impact their perceptions and opinions of CHILD, it would affect district representatives' decision to send students to CHILD. The types of mental health data in which stakeholders appear to be most interested are behavioral, (e.g., disruptiveness and/or aggressive incidents), social behavior and interaction, and self-regulation/coping skills. Other data of interest included adaptive/functional (e.g., daily life skills such as brushing teeth or preparing own meals), academic (e.g., progress on assigned curriculum, test scores, and ability to remain in the classroom), and long-term outcome data such as high school graduation and college enrollment rates, and post-secondary housing and employment status. Parents, when asked whether or not they'd be willing to share such long-term outcome data, were almost unanimously affirmative. A pre/post measurement format was most preferred amongst school district representatives.

The survey findings demonstrate a desire for more objective data across a number of dimensions, especially among school district representatives. While it is not clear if CHILD's current lack of objective data is damaging to their reputation amongst parents and school districts, it is apparent that increased data collection in certain areas would be

likely to improve consumer, or stakeholder opinion of CHILD. Therefore, CHILD would stand to benefit most from increased data collection in three areas: student mental health outcomes while at CHILD, stakeholder perceptions and opinions, and long-term student outcome data. The next two phases of the project, the macro-level literature review and micro-level dialogue with CHILD's Leadership Team will be for identifying and designing the specific instrumentation and implementation practices best-suited to meet these needs. The macro-level inquiry will consist primarily of a formal literature review to identify the best objective measure for tracking student mental health outcomes. It will also provide some context and background for tracking stakeholder perceptions and opinions and long-term student outcomes. The results from the meso- and macro-level phases will then be discussed with CHILD's Leadership Team for the purposes of refining the design of the data-tracking system.

Macro-Level Evaluation: Literature Review and System Design

Given CHILD's desire to claim expertise in the area of mental health, and based on the results of the meso-level surveys, it makes sense for CHILD to track mental health and long-term student outcome data while simultaneously keeping tabs on the perceptions and opinions of the consumers or external stakeholders of its services. This section is comprised of a macro-level analysis of each of these domains, in order to understand the relevant research and broader contextual elements influencing the final system design. Data tracking in two of the three domains will rely on custom surveys designed specifically for this purpose. The other area, student mental health outcomes while at CHILD, necessitates use of an established instrument to maximize the reliability

and validity of the acquired data. To this end, a formal literature review was conducted, the results of which are presented next.

Mental health outcome measure. The literature review for the purposes of finding an objective measure appropriate to the needs of CHILD's data-tracking system was a three-stage process. The first, or Preliminary Stage involved a review of the existing literature for meta-reviews of objective mental health measures for use with children and adolescents. Once appropriate reviews were identified, their findings were reviewed with the interests of CHILD in mind. A set of inclusion/exclusion criteria were used to identify those measures or instruments best-suited to CHILD's needs, given the results from the parent and school district representative surveys, and CHILD's own stated interests.

Preliminary stage. The objective of the Preliminary Stage of the literature review was to find comprehensive meta-reviews of mental health outcome measures, preferably those oriented towards the application of measures in schools and disorders typical of the CHILD student body population (e.g., autism). More recent reviews were prioritized, but older ones were considered if they reviewed measures still commonly used today (e.g., the BASC and the CBCL). Reviews were excluded if the domain of the review was too narrow (e.g., measures for one specific diagnosis or type of outcome), except in the case of autism, which is typically in high prevalence at CHILD. Reviews of mental health outcome measures for adults were excluded as were those lacking sufficient depth in their analysis of the measures. While there was no set cut-off for the appropriate level of 'depth', reviews that included information about the measures' purpose, psychometric properties, length or number of items, intended reporters (e.g., parents, teacher, and/or

self), and cost were sought. Six literature reviews of child and adolescent mental health measures were selected. Each review varies to some extent from the others in purpose, scope, and result, with the differences briefly described and compared below.

It should be noted the search in the Preliminary Stage produced no literature reviews oriented directly towards the measurement of mental health outcomes in a school setting. All of the reviews, as discussed below, sought measures for use in routine clinical practice, not necessarily in a school setting. If considerations were given to the use of the measures in a school setting, it was for the purposes of screening and identification of mental health issues. Interestingly, there is little to no research on mental health outcomes in a school setting. In fact, when this subject was searched across several major databases (e.g., PsycINFO, Google Scholar, OhioLink EJC) using phrases such as "mental health outcomes in school" and "mental health outcome measures for schools," only two research studies tracking progress in mental health in a school setting were found (Nickerson, Brosof, & Shapiro, 2004; Robinson & Rapport, 2002). This suggests that while there is considerable interest in the identification of mental health problems in schools (i.e., for the purposes of matching special needs students with the available and appropriate resources), there is very little in the outcomes these services produce. Given the resources invested in special education services and near ubiquitous demand for evidence of program effectiveness, this is an unusual gap in the existing research base. While the present evaluation is not intended to provide any data of this nature, the data-tracking system CHILD implements will. CHILD ultimately decides how this data will be used, shared, and presented, but if they chose to do so, presenting it for public consumption would provide a much-needed step in addressing this problematic absence

in the existing literature. It also would facilitate CHILD's growth as an innovative leader in special education, an informal but overt objective of the present Leadership Team. In order to produce data of this sort, however, CHILD must first have a reliable and valid system for collecting the data. The following section describes Stage One of the review process. It includes a description of the six meta-reviews selected during the Preliminary Stage and a list of the measures from each meta-review that met Stage One criteria for possible use in tracking mental health outcomes at CHILD.

  Stage one. Each of the reviews contains a variety of different measures, all of which were scanned and subjected to an initial selection process based on a set of broad inclusion criteria. In general, measures needed to pass basic standards of reliability and validity, be short in length (completion time of 30 minutes or less), have multiple reporters (at least two), span an age range of at least 6–16 years-old, and assess for all major domains of mental health rather than specific areas. Sound psychometric properties are important so that CHILD may use the data as valid evidence of mental health progress. A short completion time and broad age range are needed to ensure feasible utilization of the measure, and a broad scope of measured constructs is needed so that all presenting mental health issues are tracked accurately. Measures with an option for multiple reporters (at least two) were sought to provide a more robust picture of existing mental health issues and the progress students may be making in those areas. This decision is supported by research presented in the Encyclopedia of Clinical Neuropsychology (Achenbach, 2011, p. 549) that demonstrates low correlations between different informants on child psychosocial measurement tools. This suggests that multiple informants (e.g., parent, teacher, and youth) are necessary to get a more accurate and

well-rounded picture of a student's mental health issues. Finally, measures meeting criteria and appearing in multiple reviews were retained for further analysis in Stage Two.

Hunter, Higginson, and Garralda (1996). The oldest of the reviews comes from Hunter, Higginson, and Garralda (1996), who conducted a comprehensive literature review of measures for use in "routine clinical practice" (p. 198). Citing increased pressure in the mental health field to demonstrate effectiveness and value for services provided, the authors searched two databases (Medline and PsychInfo) for measures that fell within loosely defined categories. In their words, "each measure identified throughout this search was reviewed according to the scientific criteria of validity and reliability, responsiveness to change, and whether the measure can be used to evaluate child and adolescent mental health outcomes in routine clinical practice" (p. 198). As will be described later, this is similar to the process used in Stage Two of this literature review to identify specific measures with the greatest potential for effective implementation at CHILD. Furthermore, the criteria used by Hunter et al., though oriented towards use in a clinical rather than educational setting, closely align with CHILD's need to regularly track mental health progress and outcomes.

Their review spanned 75 papers, 69 of which were dedicated specifically to child and adolescent measurement tools. To organize their review, the authors identified three different types of outcomes: population outcomes, case-specific outcomes, and performance indicators. Population outcomes were defined as "changes in the health status of a population" (Hunter et al., 1996, p. 198) and included such broad, population-based statistics as homelessness and delinquency prevalence rates. The prospect of

tracking CHILD's population outcomes is worth considering in light of stakeholder interest in long-term outcome data. Parents were nearly unanimous in their willingness to share post-placement information about their child such as high school graduation, employment, and housing status, and the school district representatives made clear they would be interested in seeing it. Compiled together across time, this would provide the type of population outcome data described by Hunter et al. and could be a valuable component to CHILD's new data-tracking system.

The second type of outcome identified by Hunter et al. (1996) is the specific outcome, for which objective outcome measures are well-suited. These tools "need to be easy and quick to complete" and "should obtain information from the parent or parents, child or adolescent, and teacher or teachers" (p. 199), according to the authors. This aspect of their literature review produced a number of measures the authors believed to be well-suited for this task, based on their own selection criteria. Of those selected, four met the preliminary inclusion criteria for CHILD, meaning they require at least two different types of informants (e.g., parents and teachers, or youth self-report and teachers, etc.), can be used with a (roughly) K–12 age range, are brief in length (less than 30 min. completion time), and are broad in their scope of assessment. These were the Behaviour Problem Checklist–Revised (BPC–R), the Child Behavior Checklist (CBCL), the Conners Parent Teacher Rating Scale (CPT), and the Children's Global Assessment Scale (CGAS). Several others came close to meeting preliminary inclusion criteria, but in the authors' analysis, their ability to determine clinical change was listed as "Unknown" and so they were excluded.

The third and final type of outcome described in this review and worth mentioning here is the performance indicator. Hunter et al. (1996) define this as the type of information not directly related to service outcomes such as mental health status, but rather the daily functioning of the organization such as the "service's structure (e.g., building, equipment, staffing), processes (e.g., admission and re-admission rates, length of hospital stay, number of consultations . . .) and output (e.g., discharge rates, number of referrals)" (p. 200). These are best, according to the authors, when assessing for direct or specific outcomes is not possible. In CHILD's case, information of this sort might include the frequency of critical incidents, staff turn-over rate, students' average length of stay, and the discharge versus referral rate. While this would not necessarily provide information pertaining directly to student mental health, it could shed light on the overall health of the organization. It also provides another student outcome metric that could be used in conjunction with mental health data to illuminate program effectiveness (or lack thereof).

The review from Hunter et al. (1996) was ultimately chosen despite its age because of its comprehensiveness and thoroughness. It produced four measures meeting Stage One criteria for incorporation into CHILD's data-tracking system, which will be compiled together with qualifying measures from other meta-reviews below and subjected to a secondary filtering process in Stage Two. It also provided rationale for tracking different types of data at CHILD, and offered descriptions of the types of outcomes this might entail. Both long-term, population outcomes and objectively-measured specific outcomes would be ideal for demonstrating expertise both in mental health and in general. Performance indicators such as staff turnover and referral rates

could be useful to CHILD in its efforts to understand overall program stability, but less demonstrative of expertise in a given area.

*Levitt, Saka, Romanelli, and Hoagwood (2007).*The next review is from Levitt et al. (2007), who sought in their study to "describe the scientific status of assessment instrumentation that may be used for the range of early mental health identification strategies available to schools" (p. 164). This differs from the Hunter et al. (1996) review in two important ways. First of all, they examine mental health measures for use in a school, or educational setting. Secondly, they looked primarily at measures that can be used as screeners, or those tools designed for the purposes of identifying mental health issues rather than tracking treatment progress or outcomes on those issues. Because many mental health measures can be used as both for screening and for tracking progress, the instruments described in this review were given consideration for use as tracking instruments at CHILD.

In framing their analysis, Levitt et al. (2007) differentiate between efficacious and effective instruments. Efficacious instruments are defined as having good psychometric properties. They have been shown to have acceptable validity and reliability. Effective instruments are those most feasible for use given the context and needs of the evaluator. They can be easily, reasonably, and effectively used, in other words, within the desired setting. The length, ease of scoring and interpretation, and acceptability of the produced data are also factors contributing to the effectiveness of a tool. The authors acknowledge their list of instruments is neither exhaustive nor definitive. However, those selected were chosen "because they are commonly used both in research and clinical practice and are

likely to be efficacious (e.g., reliable and valid) as well as effective (e.g., feasible for use in schools)" (p. 169).

The review also distinguishes between three types of measures: broad, specialized, and targeted. *Broad* measures are defined as brief and straightforward assessments of mental health, usually across many areas or domains. *Specialized* measures are similar in scope, but assess in much greater detail and so tend to be longer. Finally, *targeted* measures are those assessing one specific domain of mental health, such as ADHD or depression. Measures falling in the broad and specialized categories were included in the Stage One review, while targeted measures were excluded as too narrow in scope.

Of the eight measures categorized as either broad or specialized, four met criteria for preliminary consideration for use at CHILD, and four were excluded due either to excessive length or restricted age range. Those meeting criteria included the Pediatric Symptom Checklist–35 (PSC–35), Strengths and Difficulties Questionnaire (SDQ), Child Behavior Checklist (CBCL), and Behavior Assessment System for Children (BASC), all of which have a targeted age range of at least 4–16, completion times under 30 minutes (according to the authors), acceptable validity and reliability, multiple raters, and a broad range of measured constructs. In their analysis of the psychometric properties of each measure, the authors discuss the importance of utilizing not only multiple raters in the measurement process, but the importance of including youth self-reports as well. Citing multiple research studies, they note this is especially the case for older students at the middle and high school levels, who have greater awareness and insight into their own difficulties. When students at this age level report on their own internalized symptoms

such as depression and anxiety, they tend to be more reliable than parents and teachers. On the other hand, parents and teachers tend to be more reliable reporters of externalizing symptoms and behaviors such as hyperactivity and conduct problems. All but one of the selected measures from this review, the PSC–35, has three versions: a youth self-report, a parent, and a teacher version. The PSC–35 does not have a teacher version, but it does have a youth self-report option. The downside to implementing a measure such as the PSC–35 at CHILD is that the teachers, who work daily with the students, would not have any input on the specific mental health outcomes of the students. On the other hand, this may increase to some degree the feasibility, or effectiveness of the measure because it frees the teachers up to focus on other tasks while accumulating data from other reliable sources—the students themselves (11 and over) and their parents.

Williams (2008). The third review selected for this study was arranged by Williams (2008) for the Northern California Training Academy, a social services agency concerned with the welfare of child mental health in the region. The report was generated to "offer guidance . . . in the selection of measures for universal screening and follow-up assessments within child welfare" (p. 4). Like the Levitt et al. (2007) study, a primary objective of this review was to detail screening measures. It differs, however, in that it lists outcome assessment tools available for use as well. The report organizes instruments into one of three categories: tools for assessment only, tools for screening only, and tools for screening and/or assessment. Drawing from the PsycInfo and Google Scholar databases, an iterative process was utilized in the search process with initial searches leading to new avenues for research based on various keywords and identified citations. All age groups were included. Measures selected for the review were required to assess

mental health and/or social-emotional functioning on a broad scale, while tools measuring a specific domain of mental health were excluded, along with those demonstrating questionable psychometric properties. In all, 95 total measures were identified and divided into one of the three categories: assessment only, screening only, or screening and assessment. For each measure included in the review, Williams provided a brief description of each, including the pros and cons for use in screening and/or assessment.

All measures across all three categories were reviewed for possible inclusion, based on the same criteria described above: good reliability and validity, brevity (completion time of 30 minutes or less), multiple versions for at least two different informants (at least two), a designated age range of at least 6–16 years-old, and measured constructs covering all or most major mental health domains. While screening measures are not designed primarily for tracking outcomes, they tend to be brief and therefore easier to use in a busy school setting, and research has shown that some screening measures can be effectively used as outcome measures (Williams, 2008). Measures selected here, in Stage One of the filtering process, will be subjected to an additional filtering process in Stage Two. This will be followed by an in-depth analysis of the remaining measures in Stage Three to determine the most efficacious and effective instrument for CHILD's purposes.

From the Williams (2008) review, nine measures met preliminary, Stage One inclusion criteria. Two of these were classified as assessments only, four as screening only, and three defined as screening and assessment. From the assessment only category, the Behavior and Emotional Rating Scale, 2nd edition (BERS–2) and the Social Skills

Rating System (SSRS) met criteria, and from the screening only category the Behavior Rating Profile (BRP), Child Symptom Inventory, 4th edition (CSI–4), the Strengths and Difficulties Questionnaire (SDQ), and the Pediatric Symptom Checklist (PSC) all met criteria. Finally, the three measures from the assessment and screening category that qualified are the Achenbach System for Empirically Based Assessment (ASEBA), the Behavior Assessment System for Children, 2nd edition (BASC–2), which includes its screening instrument, the Behavioral and Emotional Screening System (BESS), and the Personality Inventory for Children, 2nd edition (PIC–2).

Humphrey et al. (2011). Humphrey et al. (2011) conducted a systematic review of instruments for use in the measurement of social and emotional skills in children and young people. The purpose of the study was to provide an overview of this area of assessment and to identify and describe measures that passed their systematic review screening process. Interested in outcome measures rather than screening measures, the authors searched a number of different databases to compile an initial list of instruments. This initial list, comprised of 187 measures, was subjected to a set of inclusion/exclusion criteria. Measures not covering a broad range of social and/or emotional skills, based on professional report only (not available to parents or youth), taking longer than 30 minutes to complete, and having too narrow an age range were excluded. Those covering a broad range of social and/or emotional functioning, with an option for multiple reporters (at least two), taking less than 30 minutes to complete, and covering a broad age range were included. As this set of criteria closely resembles those of the Stage One review, there is an added degree of reliability to the list produced by Humphrey et al.

Their preliminary filtering process, once completed in detail, resulted in 23

measures. These measures were subjected to a secondary filtering process. Measures

appearing in fewer than four peer-reviewed articles were excluded. The remaining 12

measures met all of the primary criteria for breadth, length, and age range, as well as the

secondary criteria of appearing in at least four peer-reviewed journal articles. The final 12

measures were explored in depth, then compared and categorized into one of three

different categories based on the measured content of each instrument, social, emotional,

or social/emotional. While they acknowledge that "there seems to be little common

consensus as regards to what is meant by social and emotional skills, and how they are

best measured" (Humphrey et al., 2011, p. 618), they define both social and emotional

using Denham's paradigm. This model differentiates between relational/prosocial skills

such as social problem solving and relationship skills (cooperation, turn-taking, seeking

help, etc.), and emotional competence skills such as self-awareness, self-management

(emotion regulation), and social awareness. In Humphrey et al.'s view, an example of a

social measure is the Child Assertive Behaviour Scale (CABS), which measures "total

assertiveness, passivity, and aggressiveness" (p. 626, Table 4), and an example of an

emotion measure is the Emotion Regulation Checklist, which measures "negativity" and

"emotional regulation" (p. 628, Table 4). Finally, they give as an example of a

social/emotional scale: the Prosocial Tendencies Measure–Revised, which, according to

Humphrey et al., measures "public, anonymous, dire, emotional, compliant and altruistic

prosocial behavior" (p. 626, Table 4).

Because CHILD is in need of a broad-reaching measure of social and emotional

health, only those measures categorized as social/emotional were considered for

CHILD's data-tracking system. This left four measures of the original 187 reviewed by Humphrey et al. (2011). In addition to the Prosocial Tendencies Measure-Revised the other three listed in this category were the Bar-On Emotional Quotient Inventory–Youth Version, the Social Skills Improvement System (formerly the Social Skills Rating System), and the Matson Evaluation of Social Skills with Youngsters. Of these four, only the last two met preliminary criteria for inclusion to the present study. The other two were excluded based on age range (the Prosocial Tendencies Measure–Revised, with a range of 11–18) and having only a single rater (the Bar-On Emotional Quotient Inventory–Youth Version had a youth version only). The Social Skills Improvement System (SSIS) assesses 3–18 year-olds using three different versions (child, parent, and teacher), has a completion time of 10–25 minutes, and covers a broad range of social and emotional functioning. The Matson Evaluation of Social Skills with Youngsters (MESSY) assesses 4–18 year-olds using two different versions (child and teacher), has a completion time of 10–25 minutes, and also covers a broad range of social and emotional functioning. The authors also note that these two measures were the only two demonstrating sound psychometric properties at an advanced level of analysis.

Payakachat, Tilford, Kovacs, and Kuhlthau (2012). The lone review focusing on measures for autism spectrum disorders and meeting criteria for inclusion in the present study is from Payakachat, Tilford, Kovacs, and Kuhlthau (2012). The purpose of their study was to identify and review instruments available for use with "clinical, health services and cost-effectiveness applications" (p. 485) in the service of youth with autism spectrum disorder, particularly in the measurement of health outcomes. Measures were organized into three categories, clinical and behavioral measures, health-related quality-

of-life (HRQL) measures, and preference-based HRQL measures. Because the latter two focus on physical health as well as mental health, measures from these categories were excluded. This left 12 measures to analyze, as presented in the article.

According to Payakachat et al. (2012), "the outcome measures selected for this review were based on instruments that have been used in recent randomized clinical trials and/or collected in ongoing registries of children with autism" (p. 488). No further information was provided regarding the methodology of their review process. Of the 12 measures presented, only one met criteria for inclusion in the present review. This was primarily because the other measures were either too narrow in scope (e.g., focused on autism only), were too lengthy, or did not have an option for multiple reporters. The Child Behavior Checklist (CBCL) was the only measure to meet inclusion criteria for breadth of content, length, quality of psychometric properties, and number of reporters. The authors also note the CBCL "may be especially useful for measuring symptoms related to psychiatric comorbidities in children with ASDs" (p. 489).

Deighton et al. (2014). Acknowledging that previous attempts had been made to assess the psychometric properties of mental health outcome measures for youth, Deighton et al. (2014) sought to evaluate such measures not only in these terms, but also in terms of their feasibility and appropriateness for implementation in routine clinical practice. This is similar to the distinction made by Levitt et al. (2007) between efficacious and effective measures, or a given measure's psychometric properties (efficacy) and its feasibility or usability (effectiveness). They also placed particular emphasis on measures that cover a broad age range, include a child self-report option, and have sound psychometric properties. To evaluate the utility or feasibility of each measure, the authors

explored the resource implications of each, in terms of the time and money required to implement them.

The review process used by Deighton et al. (2014) consisted of four stages. The first stage was a review of various databases for tools measuring a broad range of symptoms and age ranges, with a youth self-report version, and good psychometric properties. The resource implications of each measure were also taken into account. Measures used solely for diagnostic or assessment services were excluded. On this last point, Deighton et al.'s review differs from the others in that certain measures were excluded based on the designed intent or purpose of the instrument. Thus, measures included in earlier reviews and meeting the preliminary inclusion criteria for CHILD's data tracking system (e.g., the BERS–2 and SSIS) do not appear in Deighton et al.'s review process. This omission bears consideration moving forward. Finally, the first stage included consultation with experts in the field of child and adolescent psychology and child mental health practitioners in order to gain additional knowledge of existing mental health measures. In all, this initial stage resulted in 117 identified measures.

The second stage of the process was similar to the first, but in greater depth. It also did not involve consultation with other professionals and experts. At this stage, the initial pool of identified measures was subjected to a more clearly defined list of inclusion and exclusion criteria. Instruments needed to provide a measure not only of broadband mental health but more specifically the "wellbeing in children and young people (up to age 18), including measures of wellbeing and quality of life" (Deighton et al., 2014, p. 4). Each measure was to have at least a youth self-report version, and validity properties established in a "child or adolescent context" (p. 4). Measures were excluded if

they were not available in English, focused only on a narrow or specific set of mental health domains, took longer than 30 minutes to complete, used primarily open-ended questions, covered too narrow of an age range, and had not been normed on a variety of populations. The list was reduced to 45 measures after these criteria were applied.

The third stage of Deighton et al.'s (2014) review involved a more detailed investigation of each measure individually. Particular focus was given to psychometric properties, the symptoms or content scales covered, the response format, the type of respondents or informants, the number of associated published papers, and the setting(s) in which the measure has been used. To gather this information, the authors consulted manuals, review papers, published journal articles, the test publisher when available, and other web-based resources. The only inclusion/exclusion criteria applied at this stage was that measures were excluded if no further information could be discovered about them.

The fourth and final stage of Deighton et al.'s (2014) review involved a final review of the measures for psychometric quality. In addition to the criteria established earlier, measures at this stage were subjected to a new set of criteria. Specifically, the heterogeneity of the normed samples, the extent of evidence, and the response scales were all given consideration. If the validity and reliability was only tested on one specific population or children with one type of problem or diagnosis, the measure was excluded. The depth and quality of the existing psychometric evidence was also considered. Measures included in more than five published empirical studies or researched independently by other than the measure's publisher were included; all others were removed from the list.

This robust review process identified 11 measures as most appropriate for use in routine clinical practice. All 11 of these measures also met Stage One criteria in the measure selection process for CHILD's data-tracking system. These are the Achenbach System of Empirically Based Assessment (ASEBA), Beck Youth Inventories (BYI), Behavior Assessment System for Children, 2nd ed. (BASC–2), Behavioral and Emotional Rating Scale (BERS), Child Health Questionnaire (CHQ), Child Symptom Inventories (CSI), Health of the National Outcome Scale for Children and Adolescents (HoNOSCA), Kidscreen, Pediatric Symptom Checklist (PSC), Strengths and Difficulties Questionnaire (SDQ), and Youth Outcome Questionnaire (YOQ).

The review from Deighton et al. (2014) carries significant weight in the present review process for two reasons: it is the most recent of the six reviews selected, and its review process was the most rigorous. At the same time, the authors make clear that certain measures, those intended for "assessment only," were never subjected to review in the first place. This is somewhat problematic because none of the other studies reviewed above used the intended or stated purpose of a measure as a criterion for inclusion or exclusion. It is also unclear how "assessment only" measures are defined by the authors. This makes it difficult to know exactly what types of measures may have been excluded, and whether or not those measures might have met the criteria of the others reviews discussed here. In other words, it's possible that some measures included in the other reviews were left out of Deighton et al.'s from the beginning. While this does not disqualify in anyway their results, it's worth taking into consideration.

Stage one summary. The six reviews considered as a part of the present literature review yielded 17 measures. These all met Stage One criteria for inclusion in CHILD's

data tracking system, including: minimum standards of reliability and validity, average completion time of 30 minutes or less, multiple reporters (at least two), age range of at least 6–16 years-old, and measurement of a broad range of mental health constructs. The following measures remained after Stage One of the filtering process: ASEBA/CBCL, BYI, BASC, BERS, CHQ, CSI, HoNOSCA, Kidscreen, PSC, SDQ, YOQ, MESSY, SSIS/SSRS, BRP, PIC, Conners, and CGAS. Of these, one will ultimately need to be selected for use as a part of CHILD's data-tracking system. Because Stage Two of the filtering process will require an in-depth exploration of the potential measures, this list was further pared down based on a criterion implemented by some of the studies reviewed above. As with Deighton et al. (2014), for example, the number of measures on this list was reduced based on the number of times each appeared in the reviews presented. This was considered a valid process due to the breadth and depth of the reviews, the rigor with which many of them were executed, and the fact that ubiquity of a given instrument is suggestive of an efficacious and effective instrument. Therefore, all measures appearing only once were excluded, and measures appearing in more than one review were included. These were the BASC–2, PSC/Y–PSC, and the SDQ, which appeared in three of the reviews, the ASEBA/CBCL, which appeared in five out of the six reviews, and the BERS–2 and SSIS/SSRS, which appeared in only two of the reviews, but as noted above, were left out entirely of the Deighton et al. review. It's likely that if these last two had been considered, they would have met inclusion criteria and ultimately been included in the Deighton et al. review. They would have thus appeared in three out of the six reviews, not two. These six measures are reviewed in greater detail in Stage Two.

Stage two. The final review process of the remaining six measures, the ASEBA/CBCL, BASC–2, SDQ, PSY/Y-PSC, BERS–2, involved an in-depth examination of each measure in light of CHILD's needs and the survey results of the present study. A number of different perspectives were useful in determining the types of evaluative guidelines and criteria most appropriate for finding the best measure for CHILD. Kazdin (2005) suggests "delineating the different purposes of assessment, and then, for each purpose, identifying the special requirements and then the criteria for stating when these requirements are met" (p. 548). This is useful in a broad sense when considering all components of CHILD's data-tracking system, not just the mental health outcome measure, and also highlights the importance of remembering the original purpose of the measure to be implemented. In CHILD's case, for instance, the data-tracking system will be serving several different purposes, one of which is to directly track the mental health outcomes of the students during enrollment, while the other two components are designed to track long-term student outcomes and consumer satisfaction. The details and specific components of these latter two will be discussed in a separate section below.

Establishing criteria for selecting (and developing) an appropriate measure is also a crucial part of the process. This was instituted throughout Stage One of the objective measure selection process, and will be again implemented in Stage Two. Stage Two will be somewhat different from Stage One in that it's organized around principles rather than strict cut-offs. In other words, the relevant aspects of each measure will be explored, compared in depth and considered in light of the survey data presented earlier. Rather than eliminating each measure based on a predetermined list of exclusion/inclusion

criteria, the measure that emerges as the best fit given all of the various considerations will be selected. In addition to a more in-depth exploration of the reviews included in Stage One, an additional literature review for each of the remaining six instruments was conducted, in part to find information not provided in the existing reviews and also to explore additional research on each measure.

To guide this process, it is helpful to consider McClendon et al.'s (2011) advice that "outcome measures should be brief enough to be administered on a regular basis, easily scored and interpreted, and cost effective" (pp. 111–112). These are akin to what Levitt et al. (2007) refer to as the effectiveness of a measure, or most feasible for use given the context and needs of the evaluator. These considerations and others are summarized and drawn out more thoroughly by Glover and Albers (2007), whose framework will be used to guide the Stage Two analysis and selection process. In their view, there are three aspects to consider when choosing an evaluative measure: the appropriateness for the intended use, their technical adequacy, and their usability. While theirs are recommendations for identifying universal mental health screening measures, the same principles have sufficient application to the present review. CHILD's data-tracking system is intended to be universal, for instance, and while it is not designed to screen for, or identify mental health issues, it is concerned with the use of time-efficient and cost-effective instruments typical of a universal screening process.

The appropriateness for the intended use. According to Glover and Albers (2007), a measure is "useful" if it is "appropriate for the specific administration context and selected purpose" (p. 119). They suggest four considerations when determining the appropriateness of a given measure:

- compatibility with local service delivery needs;

- alignment with constructs of interest;

- theoretical and empirical support; and

- population fit.

The first consideration, compatibility with local service delivery needs, pertains to the timing and frequency of the instrument administration, as well as the outcomes produced by the instrument. CHILD's chief concern is in demonstrating expertise in the area of mental health. Therefore, the data obtained by measuring mental health outcomes will be used primarily for the consumption of external stakeholders interested in the efficacy or effectiveness of CHILD's programming in this area. CHILD is less concerned with their day-to-day impact on a student's mental health for the purposes of improving existing programming. This is in light of the levels of internal stakeholder confidence in program efficacy or effectiveness as demonstrated in the informal survey conducted during the project's development. Thus, more frequent data collection—such as the weekly data collection typical of many outpatient settings—seems unnecessary. Instead, the frequency of measurement should be sufficient to show progress, without burdening the system of administrators and staff with additional and unnecessary documentation procedures and analysis.

The research base was consulted for insight into the ideal frequency of mental health data collection in a school setting. However, there does not, somewhat surprisingly, appear to be any current research along these lines. The reason for this is unclear, but it could be due possibly to the fact that there are no extraneous pressures on schools to present this type of data. In the IEP system as designed and legislated by the

federal Individuals with Disabilities Education Act (IDEA), student progress is measured primarily by goals and objectives drawn up in the IEP document of each individual student. While there are rules and stipulations governing this process, there is a degree of flexibility to the process that undermines it as a reliable tool of objective outcome measurement. If the IEP team reaches a consensus decision, for instance, that particular goals or objectives are outdated, unobtainable, irrelevant, or otherwise not applicable, they may be changed at any time (IDEA, 2002). The IEP document then is perhaps more accurately viewed as a localized contract between parties, subject to change as circumstances change, rather than a robust indicator of student progress.

This also underscores the importance of establishing a more robust data-tracking system at CHILD. Currently, the mental health of select students is assessed in some capacity on an annual basis using the CBCL. This yearly assessment, while perhaps somewhat arbitrarily determined, seems appropriate to CHILD's needs. Annual assessment would provide regular updates on the mental health status of the students while minimizing the amount of time required of the staff and administrators to effectively implement the system. While some measures are brief and intended for more regular use in clinical settings, this advantage may be neutralized by a lack of need in CHILD's case. CHILD currently has no direct need for daily, weekly, or even monthly assessments. On the other hand, they may at some point decide that they would like to use outcome measurement as a way of refining the quality and impact of their programming (particularly if data generated by a new data-tracking system demonstrate poor outcomes). Creating a system now that allows for this possibility down the road may be in CHILD's interest. As the emphasis and main directive of the present project is to

create a data-tracking system for the purposes of demonstrating expertise to external

stakeholders, the system should show at a minimum the overall impact of CHILD on its

students' mental health while enrolled at CHILD. A basic pre-/post- format, whereby all

students are assessed upon entrance into the program, and once more upon their exit,

would be useful. However, as many of the students are enrolled for multiple years, data

would be sparse and slow to compile. Thus, an annual assessment minimizes

administrative burden on personnel while providing a systematic and sufficient stream of

data to illustrate program-wide impact on mental health. These considerations and others

will be taken into account in the ensuring review and discussion, and reflected in the final

recommendations.

The second consideration, alignment with the constructs of interest, relates to the

mental health domains targeted by a given measure. In CHILD's instance, in wouldn't

make sense to adopt a measure for tracking mental health that is designed to assess for

attention-deficit/hyperactivity disorder (ADHD) or levels of anxiety only. Rather, an

appropriate measure would assess along a broad range of mental health dimensions,

specifically those identified by parents and school district representatives as most

relevant to their interests. The areas of greatest interest to parents and school district

representatives as suggested by the survey results reported earlier are self-regulation,

coping skills, and behavior. Stage Two of this review process will seek to clarify the

types of data generated by the different measures, so that CHILD is ensured of tracking

the data most relevant to its stakeholders. It is also worth noting here that CHILD does

have a high population of students with developmental disorders such as autism spectrum

disorder (ASD). For this reason, broad-based mental health measures with a component targeting ASD will be given special consideration.

The third aspect of a measure's "appropriateness," according to Glover and Albers (2007), is the theoretical and empirical support provided by prior research. Ideally, the measure's format and content will have been validated by previous studies, especially if these studies have incorporated use of the measure in a context similar to the one under consideration. In CHILD's case then, measures or instruments that have been used to track mental health in other similar settings (i.e., therapeutic day schools) would warrant special consideration, especially if their use was a part of a research study. In this way, the appropriateness of the measure has some empirical validation that would further justify its implementation at CHILD.

The last component to consider when determining the appropriateness of a measure for its intended use is the fit of the measure with the targeted population. As mentioned above, CHILD has a high rate of students with developmental disabilities such as ASD. They are also special needs students who have been placed outside of their home schools in a private therapeutic day school setting. The measure should not only align with the constructs of interest, but it should fit with CHILD's population in the context in which they're served. Consideration should be given specifically to the students' age range, their developmental stages, and whether or not the measure was designed to be used in schools.

Technical adequacy. The second area to consider when reviewing measures, according to Glover and Albers (2007), is the technical adequacy of an instrument. "Specifically," they write, "an instrument should be (a) appropriately standardized for use

with the target population, (b) consistent in its measurement, and (c) accurate in its identification" (p. 122). In other words, it should be normed on a population similar to the target population, and it should have good reliability and validity. Different psychometric properties will also have different relevance depending on the purpose or intended use of the instrument. For screening purposes, for instance, the validity of a measure would be essential because the goal of the measurement is to identify specific mental health issues. A measure with poor validity would render any determination made along these lines suspect at best. In CHILD's case, the primary objective is to track mental health progress across time. As such, the reliability of the instrument becomes paramount. If CHILD is dependent on a measure with low reliability to demonstrate outcomes across time, the data may not provide an accurate depiction of actual progress (or regression) in the mental health of the students.

The normative sample of the measure should also be representative of the student body population at CHILD in terms of age, gender, race/ethnicity, socioeconomic status, geographic location, and disability status. The time period and size of the sample should also be taken into account, as dated sampling from a small number of test subjects is less likely to provide a reliable norm against which to compare the measurements obtained at CHILD.

The different types of reliability of an instrument will be some of the most important psychometric properties to consider for CHILD's data-tracking system. In particular, as Glover and Albers (2007) point out, the test-retest reliability, inter-rater reliability, and internal consistency are important indicators of an instrument's ability to stand up to the potential changes brought about by time, different reporters or

respondents, and the content or constructs of a given measure. The test-retest reliability is especially important given CHILD's desire to track changes across time. Inter-rater reliability, especially with teacher/clinician versions of an instrument, is also important as it is likely that different staff members will be measuring or evaluating the same student at different intervals. Finally, the internal consistency is important to the degree that the clarity and consistency of the measured constructs are important to the evaluator.

The validity of an instrument is another way of talking about its accuracy, or the degree to which it measures what it purports to measure. Three general types or categories are described by Glover and Albers (2007), including criterion, construct, and content validity. They identify two types of criterion validity, predictive and concurrent validity. Concurrent validity is more relevant to CHILD's purposes because it is a measure of an instrument's validity as it pertains to the current status of the test subject. Predictive validity pertains to an instrument's ability to predict future difficulties, and so is more relevant to a screening instrument than a measure of progress across time. Construct validity is an indicator of an instrument's ability to measure what is says it is going to measure. This can be tested a number of different ways, including by comparison with other related measures. Content validity is similar, but pertains more to the way the different components of a measure relate to the stated purpose of the measure. Both construct and content validity are important for any instrument designed to measure mental health constructs, and in CHILD's case, especially so if they are interested in making claims of expertise to particular disabilities or areas of mental health.

Usability. The third consideration described by Glover and Albers (2007) is the usability of an instrument. While a given measure may be appropriate to its intended use and have sound psychometric properties, it may not be practical in the given context. For this reason, Glover and Albers suggest six areas for consideration when determining the usability of a measure. These include the cost of the measure, the feasibility of administration, the level of "buy in" amongst stakeholders, the available infrastructure for implementation, the availability of special accommodations if necessary, and the relevance of the obtained data to the needs of the evaluator or evaluating organization.

The cost of the measure is important perhaps for obvious reasons, but also because educational institutions in general tend to be limited in their financial capacity for adopting new procedures and systems. While the costs of the different measures are presented below, their affordability depends on the financial resources available to CHILD. Thus, the bearing cost will have on its final inclusion or exclusion in CHILD's data tracking system will depend on the micro-level (per Bronfenbrenner's, 1994, model) dialogue with CHILD's Leadership Team. If several measures appear adequate to CHILD's needs, the size of the difference in cost between the measures and CHILD's ability to financially accommodate them will have an impact on the final outcome.

The second consideration in regards to usability is the feasibility of the measure. In the view of Glover and Albers (2007), this includes the level of qualification of the test administrators, the suitability of the test formatting to the setting and target population, and the time required to administer, score, and interpret the assessment. The test should, in other words, fit with the ability of the school to accommodate it both training- and time-wise. CHILD has trained mental health clinicians on staff, including clinical

practicum students on-site that would be qualified to administer most mental health outcome measures. However, it is not necessary that these staff members are the administrators and interpreters of the evaluations, only that consideration is given to the qualifications and availability of the staff who will be responsible for the assessments. It's possible too that the different functions—administration, scoring, and interpretation—could be distributed across different staff, depending on the needs, purpose, and feasibility of the measure.

Third, the selected measure should be acceptable to the stakeholders involved in the generation and consumption of the data. The added burden of time and resources required by the new data-tracking system should be outweighed by the benefits it affords to the school and its students in the eyes of these stakeholders. In CHILD's case, this would include the Leadership Team, the staff, the students, the parents, and the school district representatives. While it can be assumed based on the survey results that the Leadership Team, the parents, and the school district representatives are all interested in objective mental health outcome data, approval of the selected measure will depend in part on approval from the Leadership Team in this regard. In other words, it will be up to the Leadership Team to determine if the selected measure will have adequate "buy-in" amongst the different stakeholders.

The fourth consideration is somewhat related to the second in that it takes into account the ability of the school to accommodate the implementation of a given data-tracking process. It differs though in that it focuses on the capacity of the school's infrastructure for adoption of the process. Rather than focusing on the qualifications and availability of the various staff, this recommendation is to consider the personnel

structure of the organization and who, specifically will manage/run the system given the design and capacity of this structure. This is where the logic model presented in Appendix A will be useful. Part of the final design of CHILD's data-tracking system will be to identify specifically the staff members who will be responsible for administrating, scoring, and interpreting the data, which will be influenced by consideration of CHILD's infrastructure.

The fifth recommendation from Glover and Albers (2007) is to ensure that the implemented measure has the necessary accommodations for use with the target population. At CHILD, some students may have a limited reading ability and so may require assistance from a staff to complete youth self-report forms. If other students have difficulty concentrating or remaining with a task for longer than a certain period of time, a longer instrument may not be suitable, especially if its length (or difficulty) would cause added stress to the students and the classroom.

Their sixth and final suggestion is to ensure the data generated by the measure is both useful and likely to improve outcomes. It's important, therefore, to understand prior to implementation how the generated data will be used. While CHILD's primary interest and rationale for implementing this system in the first place is to begin to demonstrate expertise in the area of mental health, some questions remain as the final utility of the data. How, for instance, will CHILD respond to a student that is not improving based on the objective data? What if the progress demonstrated by the data is negligible, or statistically insignificant? What if the data shows that, on average, student mental health is actually regressing across time at CHILD? Will CHILD use the data to make adjustments to its programming, so that mental health outcomes are more favorable?

While some of these questions are secondary to the stated purpose of the data-tracking system, they are worth noting given the nature of the information the data-tracking system will be designed to produce.

Measures. The measures appearing first will be those appearing least frequently in the meta-reviews above, the BERS–2 and SSIS, both of which appeared in two out of the six meta-reviews. These two will be followed by the PSC, SDQ, and BASC–2, which appeared in three of the meta-reviews. Finally, the CBCL, which appeared in five out of the six meta-reviews will be discussed last. Each measure is afforded its own section in order to thoroughly describe and assess it in relation to CHILD. The recommendations from Glover and Albers (2007) are used to guide this process. The section concludes with a summary synthesizing the presented material and is followed by a series of formal recommendations for CHILD's data-tracking system.

Behavioral and Emotional Rating Scale–2nd Ed. The Behavioral and Emotional Rating Scale–2nd Ed., or BERS–2, is a "multi-modal assessment system that measures . . . several aspects of a child's strength" and is "designed for use in schools" (PRO–ED, Inc., 2012). There are three versions: parent, teacher, and youth self-report (Williams, 2008), and is, according to the test-maker, designed to be used as a screener and as an outcome measure (PRO–ED, Inc., 2012). It is one of the shorter instruments under consideration, with 52 items (Deighton et al., 2014) and has an average completion time of 10 minutes across all three versions. It is has been normed on both clinical and non-clinical samples and is appropriate for youth aged 5–18 (Williams, 2008). Its content scales specifically measure interpersonal strength, involvement with family, intrapersonal strength, school functioning, affective strength, and career strength (Buckley & Epstein,

2004), the same five factors appearing in the original BERS (Buckley, Ryser, Reid, & Epstein, 2006).

The validity of these constructs and reliability of the test has been fairly well established across a number of different studies. According to Buckley and Epstein (2004), content validity was established by having parents and professionals rate children with and without emotional/behavioral disorders based on objective statements about emotional and behavioral strengths. Results from this process were then incorporated into the test items. Deighton et al. (2014) reported the teacher and parent report versions can discriminate between a sample with and a sample without emotional and behavioral problems, and the teacher version is also able to discriminate between students with and without learning disabilities. Buckley and Epstein note that it has high convergent validity with the SSIS–RS (discussed below) and a moderately negative correlation with the Achenbach Problem scales (to be expected given its design as a strength-based measure). Deighton et al. also report high internal consistency and test-retest reliability across all versions of the instrument (0.79–0.98 and 0.82–0.99, respectively). One week test-retest reliability was also high (over 0.80), and repeated studies have shown both short-term and long-term (over six months) stability across time. All three versions of the test have been normed on a nationally representative sample both with and without presenting mental health disorders (Buckley & Epstein, 2004).

The BERS–2 costs $198.00 for the starter kit, which includes the manual, 25 of each test version, and 50 summary forms. Additional test versions and summary forms cost $37.00 per package (25/test version and 50/summary form) (PRO–ED, Inc., 2012). Based on CHILD's current enrollment of almost 50 students, it would cost them $309.00

for the first year of use. This is assuming annual measurements. Every subsequent

school-wide measurement would cost $259.00, assuming six packages of 25 test versions

(two each of the parent, teacher, and student forms) and one package of 50 summary

forms at $37.00 each. As the measure is designed to be used in schools, it could easily be

scored by non-mental health and mental health specialists alike. The brevity of the

measure would lessen administrative burden on CHILD's staff, and allow for more

simplified administration to the students given their often limited attention span and

difficulty adjusting to changes in routine.

The BERS–2 in many ways would work well in light of CHILD's data tracking

needs. It's compatible with local service delivery needs in that it is brief, intended as an

outcome measure (amongst other uses) for use in schools (amongst other settings), and

available for frequent administration. It fits with CHILD's target age range. It has sound

psychometric properties, and rates well in its usability: it's affordable relative to other

measures reviewed below and could be feasibly administered in an environment such as

CHILD's. However, the BERS–2 is a strength-based measure, designed to measure inter-

and intrapersonal strength, involvement with family, affective strength, and school

functioning. These constructs do not appear at face value to overlap clearly with those

identified by parents and school district representatives as most appealing—student

behavior, coping skills, and ability to self-regulate. On the other hand, it's possible that

inter- and intrapersonal strength could represent a student's self-regulation and coping

skills, for instance, or that school functioning could act as an indicator of student

behavior. The remaining constructs, though—involvement with family and affective

strength—would not have a clear application. In fact, both parents and school district

representatives showed the least amount of interest in metrics related to emotions or affect. In summary, it's possible that BERS–2 would work well at CHILD, but may not be the ideal measure due to its lack of alignment with the mental health constructs of interest.

Social Skills Improvement System Rating Scales. The Social Skills Improvement System Rating Scales, or SSIS–RS, is a revision of the Social Skills Rating System (SSRS) and debuted in 2007 (Gresham & Elliott, 2007). It is perhaps due to its relatively recent, and, according to Gresham and Elliott, substantial revision that the SSIS–RS is somewhat less prominent in the literature than the other measures presented in this review. Its novelty should not necessarily rule it out for use at CHILD, however. The SSIS–RS has a number of factors that make it a candidate for inclusion in CHILD's data-tracking system. It is designed for all ages served at CHILD (age range is 3–18 years), has three versions (parent, teacher, and youth), can be used as an outcome measure, and is designed specifically for use in schools (Williams, 2008). According to Crosby (2011), it "provides assessment information on [positive behaviors] as well as information on problem behaviors that may interfere with the student's ability to acquire or perform specific social skills" (p. 292). The focus is on both positive and negative outcomes— unlike the strengths-based focus of the BERS–2—across three domains: social skills, problem behaviors, and academic competency.

The test itself has between 75 and 83 items, depending on the version, and has an average completion time of 10–25 minutes (Humphrey et al., 2011). This makes it a longer test than the BERS–2. Possible responses are arranged on a four-item Likert scale, from "never" to "almost always" on the parent and teacher versions, and "not true" to

"very true" on the youth self-reports. For students aged 13–18, responses can be provided regarding the relative importance of the social skills or problem behaviors. One advantage to the SSIS–RS is its inclusion of an autism spectrum scale. This differentiates it from the BERS–2, and is suggestive of a better population fit at CHILD. The various scoring forms can also be combined into one aggregate report, called an AIR form. According to Crosby (2011), "an examination of the SSIS–RS will leave one with the impression that these are indeed tools for measurement within a social/behavioral model of student functioning" (p. 295). While this does not confirm it as a measure of the specific domains preferred by CHILD's stakeholders (behavior, coping skills, and self-regulation), it is evidence that it would work well as a broad-based measure of mental health outcomes.

Psychometrically, the SSIS–RS seems to perform at a satisfactory level, if slightly lower than the BERS–2. According to Humphrey et al. (2011), the measure "correlates with other similar measures" (p. 630) and has high discriminant validity in ADHD evaluation. Crosby (2011) reports that test scores between individuals with and without social/emotional and/or communication impairment were statistically significantly different. Tests to determine concurrent validity with the BASC–2 and Vineland Adaptive Behavior Scales (2nd ed.) were also "encouraging" (Crosby, p. 295). Specific test items were constructed based on individual expertise and key items/terms from the *DSM-IV-TR* and its reliability across the test's different content scales ranges from 0.70 to 0.80 and above. Correlations between test versions demonstrated fair consistency across the three main content scales. Finally, Humphrey et al. reported the test's internal consistency across all three versions ranged from 0.73–0.97, its test-retest reliability

ranged from 0.59–0.81 on the youth version to 0.72–0.87 on the parent version, and inter-rater reliability ranged from 0.38–0.69 on the parent version to 0.48–0.69 on the teacher version. The inter-rater reliability scores are not sufficiently low to exclude the SSIS–RS from consideration for CHILD, but given the premium CHILD is placing on tracking progress over time, these numbers are worth considering. Since different teachers and possibly different parents would be evaluating the same student over time, lower inter-rater reliability could weaken the evidence of objective progress.

The final consideration of a measure, in Glover and Alber's (2007) view, is its usability in the setting for which it is intended. This includes several factors such as cost, feasibility of administration, the available infrastructure for implementation, and the usefulness or relevance of the obtained data in light of the costs associated with its use. Relative to the BERS–2, the SSIS–RS is somewhat more expensive. A starter kit with 3 packages of 25 hand-scored tests costs $271.50, with additional packages of 25 tests costing $58.65 each (Pearson Education, Inc., 2015b). This means that at a minimum, it would cost CHILD $447.45 for its first year of use, and $351.90 annually thereafter, based on annual test administration. For twice yearly test administration, the annual cost would double to $703.80. One advantage to the SSIS–RS is the availability of a computer-based scoring system, which would reduce administrative burden on CHILD's staff and administration in implementation of regular evaluation. Annual costs would remain the same with the software, but the initial startup cost would increase by $294.05, bringing CHILD's initial investment for one year of test administration to $741.50 (Pearson Education, Inc., 2015b). This is more than double the cost of the BERS–2, but affords the added ease of computer scoring and tracking.

It's also important to consider the feasibility of administering the SSIS–RS on a regular basis at CHILD. This includes an analysis of the time and resources required to implement it effectively. While the test itself is somewhat longer than the BERS–2 in terms of average completion time—over twice as long by some estimates (Humphrey et al., 2011; Williams, 2008)—a computer-based scoring system would theoretically reduce the amount of time needed by staff to compile, store, and analyze the data. On the other hand, the SSIS–RS requires mental health practitioners to score and interpret the test (Crosby, 2011). This is not the case for the BERS–2 and other instruments described below. This does not immediately exclude the SSIR–RS from consideration, but is another factor to consider when analyzing its feasibility. Will CHILD have the resources amongst its limited mental health staff to effectively implement this instrument? The other major consideration is whether the mental health constructs targeted by the SSIS–RS align with CHILD's needs. Its three main content scales measuring social skills, problem behaviors, and academic competence may have somewhat more relevance than the strength-based factors of the BERS–2.

Of greater importance, however, is their overlap with stakeholder interest in coping skills, self-regulation, and behavior. Does the measure align with the constructs of interest? Each content scale on the SSIS–RS is broken down into a number of subscales. The social skills scale is comprised of communication, cooperation, assertion, responsibility, empathy, engagement, and self-control; the problem behaviors scale, which measures behaviors that interfere with acquisition or performance of socially appropriate behaviors, includes the subscales of externalizing, internalizing, hyperactivity/inattention, autism spectrum, and bullying; and the academic competence

scale measures ability in areas of reading, math, motivation, parent support, and general impressions of cognitive functioning (Crosby, 2011). It also includes an autism spectrum subscale, suggesting good population fit with CHILD. Considering these scales then, the SSIS–RS does seem to have components that would overlap with self-regulation (e.g., the self-control subscale) and behavior (e.g., the problem behavior content scale), but less so the domain of coping skills, at least at face value. Nonetheless, it is a broad measure of mental health with a good population fit at CHILD that is likely, for the most part, to obtain relevant data. The exception to this is the academic competency scale, which is beyond CHILD's needs at the moment.

In summary, the SSIS–RS appears to have some potential use at CHILD, particularly in its comprehensiveness, relatively good alignment with CHILD's constructs of interest, and population fit. However, it is more expensive than the BERS–2, especially with the computer-based scoring system, it takes longer to administer, its psychometric properties, while sufficient, may be somewhat lower in one or two key dimensions, and it requires a mental health specialist to score and interpret the data. Its greatest asset in relation to CHILD is probably its fit with the student body population and general alignment with constructs of interest (although not perfect), but its usability may be low due to its scoring protocol and cost.

Pediatric Symptom Checklist. The next measure under consideration is the Pediatric Symptom Checklist, or PSC. Referred to as a "broad instrument" (Levitt et al., 2007, p. 173) for screening behavior, social, and emotional problems (Levitt et al., 2007; Williams, 2008), it was designed originally to be completed by parents in the waiting rooms of physicians', or pediatricians' offices (Jellinek et al., 1988). Shorter than both

the BERS–2 and the SSIS–RS, it has an average completion time of 5–10 minutes (Levitt et al., 2007). Its 35 items (Deighton et al., 2014) are designed to yield a single score of "general overall psychosocial functioning" (Jellinek et al., 1988, p. 201) that can be used to detect for presence of mental health conditions. While originally designed for use in a medical setting, its application has broadened to include school settings (Gall, Pagano, Desmond, Perrin, & Murphy 2000; Levitt et al., 2007), and is still primarily used as a screening measure (Jellinek et al., 1999; Levitt et al., 2007; Williams, 2008). The test is not designed for children older than age 16 or younger than age four, and has two versions, a parent and youth self-report. The youth self-report is for children aged 11 to 16, meaning that students at CHILD under the age of 11 would only have one reporter on their mental health status, a parent. However, it is still under consideration for use at CHILD because it met all preliminary criteria, appeared in three out of the six meta-reviews, and has several advantages detailed below.

One of these advantages is the cost. The PSC is a publicly available instrument, readily available online at no fee to the user. While CHILD may have a budget to incorporate other, more expensive measures, they may be interested in cheaper alternatives, particularly if they are sufficient for the present purposes. The other advantage to the PSC is its brevity and ease of use, which would make it easy not only to administer, but also to interpret and utilize effectively as well. It is designed to produce a single score reflecting overall mental health, and so does not necessarily require oversight from a specialist or mental health expert. In a budget-strapped, administratively-burdened school environment, inexpensive, easy-to-use tools may be most preferred.

Psychometric properties of the PSC have proven to be sufficient in most areas. It has shown to discriminate well between referred and non-referred children, both with and without mental health problems. Research has demonstrated that concurrent validity between the parent version and the CGAS and the CBCL is moderate to high at 0.79–0.92 and 0.52, respectively. The youth self-report has demonstrated a lightly lower validity relative to other similar measures, however, with scores ranging from 0.42–0.58. In general, the reliability is adequate, with internal consistency and one-week test-retest reliability on the parent version at 0.89 and 0.86, respectively. Test-retest reliability on the youth self-report version was lower at 0.45, but this was at the four-month, not one-week mark (Deighton et al., 2014). Navon, Nelson, Pagano, and Murphy (2001) demonstrated a four-six week test-retest reliability of 0.80.

Limitations already mentioned include its limited number of reporters (only two, a parent and youth self-report version, compared to three with the BERS–2, SSIS–RS, and others) and the fact that the youth self-report could only be used with about half of CHILD's student body population due to its age range. It was also designed to be and is still primarily used as a screening instrument rather than an outcome measure of mental health. If CHILD were to take a greater interest in specific aspects of its students' (or a single student's) functioning, the PSC would not be able to provide such in-depth information. The PSC does not have content scales like the BERS–2, SSIS–RS, and others do. It produces only a single score, an indicator of the subject's "psychosocial functioning" at that time. While this may incorporate at a general level aspects of functioning pertaining to coping skills, behavior, and self-regulation, CHILD would have no way of claiming progress in these areas specifically, possibly undermining their ability

to make claims to expertise in the areas of mental health most relevant to their

stakeholders.

There are three other noteworthy limitations to the PSC. The first is its three-point

Likert format to the test item responses. As Deighton et al. (2014) point out, tests with

three response options instead of four or more, as are found on many of the other

measures, are more susceptible to floor/ceiling effects and likely to be more insensitive to

change across time. Another potential weakness is that while PSC test sensitivity is high

(0.95), its specificity is lower (0.68) (Jellinek et al., 1988). This means, essentially, that if

the results are negative, the high sensitivity indicates it is very likely an accurate

indication; but if a student tests positive for mental health issues, there is a 32% chance of

a false positive. As a screening measure being considered for the purposes of outcome

measurement, this is somewhat problematic. However, as a measurement of progress

across time, a one in three chance that positive test results are false would substantially

undermine CHILD's ability to make claims to expertise in that area. Finally, the PSC

does not address any issues related to autism spectrum disorder, and so does not fit as

well with CHILD's population as say, the SSIS–RS does.

In summary, it appears that while the PSC offers some clear advantages in cost

and brevity, these may be offset by response-set limitations in the number of informants,

potential for floor/ceiling effects, and insensitivity to change across time. Relatively

inadequate population fit, problems with specificity, and questionable test-retest

reliability further undermine the case for adoption of the PSC at CHILD.

Social Difficulties Questionnaire. The fourth instrument for review is the Social

Difficulties Questionnaire, or SDQ. It met all preliminary inclusion criteria through Stage

One of this review process, and appeared in three of the six meta-reviews explored for this project. It is intended as a screening measure of strengths and problems, including emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior (Williams, 2008). Though it was designed originally to be used as a "behavioral screening measure" (youthinmind, 2012), it has since been declared suitable for use "as a part of a clinical assessment, as a treatment-outcome measure, and as a research tool" (Goodman, 2001, p. 1337). While it has a number of similarities to the PSC and the other measures discussed above, it differs in some key areas.

At 25 items, it is the shortest of the four presented thus far, and has an average completion time of about five minutes (Williams, 2008). It has various versions appropriate for any youth aged 2–17 (youthinmind, 2012) and is considered by its publisher and others in the field to be a broad (if brief) instrument of psychosocial evaluation (Goodman, 2001; Levitt et al., 2007; Williams, 2008). Like the PSC, responses are marked on a three-point Likert scale ranging from "Not true" to "Certainly True" which introduces increased risk for floor/ceiling effects and decreased sensitivity to change across time. While there are different versions for teachers, parents, and youth, as well as younger versus older students/children, the versions vary only slightly to reflect age-related changes in preferences and language.

Each version has five content scales, comprised of five questions each, and reflecting a mix of positive and negative traits. The hyperactivity-inattention scale evaluates whether a student is distractible, persistent, restless, fidgety, and reflective. The emotional symptoms scale assesses fears, worries, clinginess, unhappiness, and somatic symptomology. The prosocial behavior scale asks whether a child helps out, shares, and

is caring, considerate, and kind to others. The conduct problems scale includes questions related to lying, fighting, temper, stealing, and obedience. Finally, the peer problems scale has questions asking whether the child has a good friend, is popular with others, behaves better with adults, tends to act or play alone, and is a victim of bullying behavior. Each scale, with the exception of the prosocial scale, can be summed to provide a "total difficulties" score, but unlike the PSC, different content scale scores can be combined to create other indicators. The conduct and hyperactivity-inattention scales combined together, for instance, produce a score reflecting a child's externalizing behaviors and the emotional and peer problems scales when combined together form a score of internalizing behaviors (Goodman, 2001). These constructs, while designed to be comprehensive and broad-reaching, nonetheless fail to incorporate symptoms or behaviors related to autism spectrum disorders. It is also not clear at face value whether constructs considered most relevant by CHILD's stakeholders–coping skills, behavior, and self-regulation–are covered by the SDQ. There are some questions regarding behavior (for example, in the prosocial behavior and conduct problems scales), but these behaviors relate more to an emotionally/behaviorally disabled population than an autism spectrum demographic. While there is certainly overlap between these two at CHILD, this analysis suggests only a moderate population fit, and less ideal than others presented above, such as the SSIS–RS.

The psychometric properties of the SDQ are generally good, with a few exceptions. In terms of validity, it seems to have good discriminatory power between clinical and normative populations (Deighton et al., 2014). Goodman (2001) also compared the SDQ with the CBCL (discussed below), which is a longer, more involved

measure than the SDQ. High concurrent validity was demonstrated between the two measures, and it was shown the externalizing and internalizing scales correlated less on the SDQ than the CBCL, indicating greater discriminatory power between these two constructs on the SDQ. The SDQ was also shown to have greater sensitivity to hyperactivity-inattention. However, this is less relevant to CHILD from an outcomes standpoint.

The reliability on the SDQ is somewhat variable, according to much of the research. While its internal consistency across all versions is good (0.63–0.88), its test-retest reliability is lower. With four to six months in-between tests, reliability on the teacher and parent versions were in the moderate to good range (0.65–0.82 and 0.57–0.72 respectively), but youth self-report version ranged from 0.21 to 0.62 (Deighton et al., 2014). Finally, the mean cross informant correlation, or the degree to which different informants' responses about the same child correlated, was in the low to moderate range at 0.34 This last number is not necessarily an indicator of poor psychometric, however. Instead, it is argument for multiple informants on a given measure because if different informants have different perspectives about the same child, these should be included in the final evaluation of that child or student.

As noted above, the SDQ is a very brief, easily administered, scored, and interpreted measure. It does not require, as the SSIS–RS does, mental health expertise in its analysis. In addition to these qualities, the SDQ, like the PSC is also free. These aspects of its usability comprise the greatest strengths of the SDQ in relation to its potential use at CHILD. However, some questions remain about its goodness of fit to the student body population, and whether the data it produces will be most relevant and

useful to stakeholders, both internal and external. The SDQ certainly may track mental health data, and it may do so with some accuracy. However, if this data is of little interest to the parents and school district representatives that work with CHILD on a regular basis, than there is little point in implementing a measure such as the SDQ.

Behavior Assessment System for Children–2nd ed. The next measure under consideration is the Behavior Assessment System for Children–2nd ed., or BASC–2. According to its publisher, the BASC–2 is a multi-method, coordinated set of tools that can be used for mental health evaluation, diagnosis, and/or intervention planning (Pearson Education, Inc., 2015a). As such, it is simply larger and more "sophisticated" (Reynolds, Kamphaus, & Vannest, 2011, p. 367) than the other four measures presented thus far both in scope and comprehensiveness. The literature review process revealed the BASC–2 has an abbreviated version called the Behavior and Emotion Screening Scale (BESS) that can be used in conjunction with and as a part of the BASC–2. As it states in its name, this shorter test is designed to be used as a screening tool for mental health issues. It is also a way of summarizing and monitoring the mental health status of groups, organizations, and schools, which is precisely what CHILD seeks in its data-tracking system. While this may give it certain advantages over the other tools already described, there are a number of considerations to take into account before recommending the BASC–2 and BESS to CHILD as an ideal data-tracking tool.

One area to consider, as suggested by Glover and Albers (2007) is the appropriateness of the measure for the intended use. Is the BASC–2 (and BESS) compatible with CHILD's service delivery needs? Does it align with the constructs of interest and fit with their student body population? Has there been substantial theoretical

and empirical support? These questions are important, and the answers complicated by the existence of the BESS, which can be treated, like all of the components of the BASC–2, as an independent measure. Much of the research on the BESS, in fact, has regarded it as such (DiStefano & Kamphaus, 2007; Dowdy, Furlong, Eklund, Saeki, & Ritchey, 2010; Kamphaus et al., 2007; Renshaw et al., 2009; Wallbrown, 2013) and one of the meta-reviews described earlier (Williams, 2008) treated the BESS and the BASC–2 as separate instruments. However, the items on the BESS stem directly from the BASC–2 (Renshaw et al., 2009), and according to the test-maker, the BESS is designed to be used as the first stage in a three stage process involving screening, targeted or focused assessment, and comprehensive, diagnostic assessment (Pearson Education, Inc., 2015a). It is through this model of intended use that the BASC–2 will be explored, with the BESS treated as component part to the larger BASC–2 system.

The BASC–2 has a number of different components, in addition to the BESS, that can be used individually or in combination, depending on the needs of the evaluator. These include a structured developmental history (SDH), a student/portable observation system (SOS/POP), a parenting relationship questionnaire (PRQ), teacher and parent rating scales (TRS and PRS), a self-report of personality (SRP), and a self-report of personality interview (SRP-I) for younger children ages six to seven. Each of these vary in terms of content and purpose; the most relevant to CHILD's purposes, in addition to the BESS, are the TRS, PRS, SRP, and SRP-I. In essence, the BASC–2 has three informants at each stage of the assessment process: teacher, parent, and youth self-report. The BASC–2 has the added advantage of a youth self-report version for children as young as six years old (Pearson Education, Inc., 2015a). This is four to five years

younger than the self-report versions of other tests (which typically begin at age 10 or 11) and essentially would allow for input from three informants at every age/grade level at CHILD. The coverage would also act as a failsafe for students whose reading levels were below grade average and in need of an interview format.

How does the BASC–2 align with the constructs of interest to CHILD's new data-tracking system? It can be used in virtually any environment where mental health is emphasized, including mental health and pediatric clinics, community programs, and schools (Williams, 2008). This is due in part to the fact that its items and content were organized primarily around the diagnostic definitions of the *DSM-IV-TR* and legislative mandates of IDEA (Pearson Education, Inc., 2015a). As explained by Reynolds et al. (2011), the BASC–2 can "facilitate the differential diagnosis and educational classification of a variety of emotional and behavioral disorders of children and . . . aid in the design of treatment plans" (p. 366). But are the constructs it measures of any use or interest to CHILD and its stakeholders? The test has four main content or composite scales: adaptive skills, behavioral symptoms, externalizing problems, and internalizing problems, although each test version (the TRS, PRS, and SRP) varies somewhat in its organization and composition of different characteristics. This is most evident in the SRP which dispenses with the "externalizing problems" and "adaptive skills" categories altogether in favor of an "emotional symptoms index" and a "personal adjustment" scale while adding a "school problems" composite. The composite scales on this version of the BASC–2 (the SRP) are composed of the following sub-scales: anxiety, depression, self-esteem, self-reliance, sense of adequacy, and social stress (emotional symptoms index); attention problems and hyperactivity (inattention/hyperactivity); anxiety, atypicality,

depression, locus of control, social stress, sense of inadequacy, and somatization (internalizing problems); interpersonal relations, relations with parents, self-esteem, and self-reliance (personal adjustment); and attitude to school, attitude to teachers, and sensation seeking (school problems). The main indices on the TRS and PRS are composed of the following dimensions: activities of daily living (PRS only), adaptability, functional communication, leadership, social skills, and study skills (TRS-child and adolescent forms only) (adaptive skills); aggression, attention problems, atypicality, depression, and withdrawal (behavioral symptoms); hyperactivity, aggression, and conduct problems (child and adolescent forms only) (externalizing problems); and anxiety, depression, and somatization (internalizing problems) (Pearson Education Inc., 2015a). The BASC–2, judging by its composite scales and various dimensions of assessment may be the most comprehensive of the measures explored thus far, both in breadth and depth. Although it does not overtly measure two of the primary constructs of interest to CHILD's stakeholders, coping and self-regulation skills, it could be argued that these dimensions are accounted at least in part by the adaptive skills composite. At the very least, it is robust in its evaluation of the other primary construct of interest, behavior, and relative to the other measures would seem to provide a more substantial profile of a student's mental health. Given CHILD's primary interest in tracking population, rather than individual health, however, the collection and interpretation of a detailed profile of each individual student's progress may be excessive and unnecessarily burdensome to staff and the overall CHILD system, especially when considering the time and resources required to collect, store, and interpret this amount of data.

This is where the appropriateness of a given measure overlaps with Glover and Albers's (2007) description of a measure's usability, specifically its feasibility of administration. With between 100 and 185 items, the TRS, PRS, and SRP of the BASC–2 have an average completion time of 25–30 minutes (Williams, 2008) or even longer on the SRP (Levitt et al., 2007). While this did not disqualify them from consideration at CHILD during Stage One of this review, it bears reconsideration against the shorter, arguably more nimble measures such as the PSC and SDQ. This is also where the evaluator's initial internal role has added benefit to the project: it was clear working within the organization that staff members would appreciate the adoption of shorter tools; time always seemed to be at a premium. As a screening measure with 25–30 items, therefore, the BESS (Renshaw et al., 2009) supplies a certain brevity and ease of administration that the other components of the BASC–2 do not. It could also allow for more frequent administration if CHILD so chose. Thus, the potential burden of the BASC–2 on limited staff resources could be offset by strategic implementation of the BESS, either in conjunction with the TRS, PRS, and SRP or otherwise.

While the BASC–2 appears more or less appropriate for its intended use at CHILD, additional inquiry into its usability and technical adequacy may clarify its potential for use at CHILD. First off, because this relates to its appropriateness for intended use discussed above, it should be noted that an emerging body of research is demonstrating high discriminatory validity with the BASC–2 in its assessment of autism spectrum disorder (Mahan, & Matson, 2011; Volker et al., 2010). Although CHILD wouldn't be in a screening position for ASD, its sensitivity to features of the disorder could be a useful component in the data-tracking system. Deighton et al. (2014) reports

that the BASC–2 also has high discriminatory validity in general across its different clinical profiles, reaffirming its goodness of fit with CHILD's student body. Its concurrent validity with other similar measures, however, is somewhat more variable. The PRS version correlates highly with the CBCL (0.71–0.84) and the SRP with the MMPI-2 (0.78–0.89), but the TRS was found to correlate with the SSIS–RS in the low to moderate range (0.03–0.6). Reliability on the BASC–2 also has tested high across different domains. Internal consistency tends to range from 0.74–0.90, one-month test-retest reliability from 0.64–.096, and internal consistency across gender from 0.64–0.90 (Deighton et al., 2014). Importantly, scores on the BESS scales specifically have been shown to correlate highly and at a statistically significant level with the composite scales of the TRS, PRS, and SRP (Wallbrown, 2013). This lends credibility to the use of BESS in conjunction with the other components of the BASC–2.

A possible drawback of the BASC–2 is its length. The information the BASC–2 is capable of providing about its subjects may simply be more than what is needed. However, the availability of the abbreviated BESS could provide CHILD with a shorter form to be used in conjunction with the longer BASC–2 components, the TRS, PRS, and SRP. Together, these components could be implemented strategically to increase usability and maximize both the quantity and quality of the collected data. There are two additional areas of investigation: the available infrastructure of the organization—who, specifically would administer, score, and interpret, and how would they do it?—and the cost of the measure. The first of these two questions will depend in part on the final phase of this project's data collection, the micro-level dialogue with CHILD's Leadership Team. The school continues to change across time, certainly since this project first began,

and so the specific details of implementation will need to be ironed out in dialogue with the Leadership Team. At the same time, there are aspects of the BASC–2 that will dictate the way CHILD uses it. For example, the length of the TRS and SRP would necessitate special planning on the part of the teachers to make room for their administration during the school day, certainly more so than the much shorter BESS. Then, once these are administered, provisions must be made for the entry and interpretation of the data. Who, specifically, will compile the data, and what will be done with it once it is? Unlike the SSIS–RS, the BASC–2 does not require a mental health specialist for this latter phase. Anyone can score and interpret the data, which is part of what makes it readily available to so many different contexts and environments. What should be considered, however, is the quantity and complexity of the data that is likely to be generated by a tool as comprehensive and involved as the BASC–2. One compelling aspect to the shorter, simpler measures such as the BERS–2 and others described above is not only their ease of implementation, but the simplicity of the data generated. The PSC produces exactly one number. So does the BESS (Pearson Education, Inc., 2015a). However, as noted above, there are limitations to such brief, simple measures. The key is in the balance between the two.

One idea would be to use the BESS in conjunction with the TRS, PRS, and SRP. The longer versions could serve well in a sort of pre-/post- format, with the BESS acting as an intermediary progress tracker over the course of the intervention, in this case, the student's enrollment at CHILD. In other words, every student could have a TRS, PRS, and SRP completed once upon entry and once again upon graduation or transfer from CHILD. This would provide a more global, longitudinal look at the student's mental

health progress. In the meantime, as a way of tracking of progress more "locally" and frequently, the BESS could be used at regular intervals to provide a quick snapshot of each student as they progress through CHILD. This would serve a dual purpose. First, by creating more data from a greater number of sources on each student, the statistical power of the data collected would be increased and thus any claims to expertise CHILD would like to make based on that data would be strengthened. Second, the increased frequency of the data would allow CHILD to have a more up-to-date understanding of their students' mental health, and while this is not the primary intent of the proposed data-tracking system, they could use this data to help tailor programming to the needs of each student as they progress through CHILD. The frequency of implementation of any version of the BASC–2 will depend in part on cost, and the results of dialogue with CHILD's Leadership Team, based on information provided to them. Relative to the other instruments presented thus far, it is somewhat more expensive, but perhaps not significantly so.

While there are a number of different cost options for the BASC–2, depending on the types of tests CHILD would use, how frequently they would use them, and whether they would use hand or computer scoring, the option these makes the most sense is a web-based administration, scoring, and reporting system that costs $125.50 to purchase outright, and then an additional $2.40 per test administration and report for the basic score summary. Costs go up from depending on how much detail CHILD wants on each report. For CHILD's purposes at this point, the basic summary makes the most sense (however, they would be able to pay for more detailed reports if they were interested later). Therefore, to administer the BASC–2 once annually to a student, one of their

parents, and a teacher, it would cost approximately $360 (based on 50 student, teacher, and parent reports, per year, at $2.40 each). Combined with the $125.50 start-up cost, it would be around $500 for the first year at CHILD, and about $360 each subsequent year. These are estimates, but use of the TRS, PRS, and SRP then would be slightly more expensive than the BERS–2, and less expensive than the SSRS-IS. Unfortunately, it does not appear that BESS is being carried yet on the same web-based platform as the TRS, PRS, and SRP; it is only available as a pencil-and-paper test. Therefore, there would be additional costs to incorporating the BESS, even though it is a part of the BASC–2 system. A BESS start-up kit, which includes the manual and 25 each of the parent, teacher, and student forms, costs $136.50. Additional packages of 25 are $28.30 each. Therefore, for the first year of administration of the BESS, it would cost CHILD approximately $221.40 based on 50 students and a single administration. In subsequent years it would cost CHILD $169.80 based on the same number and frequency of test administrations. If CHILD was interested in using the test-maker's computer-assisted scoring software, this would cost an additional one-time fee of $667.95 for the software. Combined with the web-based version of the TRS, PRS, and SRP, without the computer-assisted software of the BESS, it would cost CHILD approximately $721 for the first year, and $529 annually thereafter. If they were to purchase the computer-assisted scoring software for the BESS, their costs would increase to $1388.95 for the first year, and remain at $529 thereafter. This is significantly more expensive than both the BERS–2 and SSIS–RS.

As with the other measures, there are pros and cons to the BASC–2. It is more comprehensive, and capable of providing considerably more data on each student.

CHILD would have options of expanding their use of the test depending on their needs as they change across time. It has both long and short forms which could be used in tandem to create a more in-depth profile of each student's progress while at the same time satisfying CHILD's needs to amass aggregate data on student body mental health for the purposes of demonstrating expertise. The BASC–2, above all else, provides CHILD with options. On the other hand, it also presents some challenges. Logistically speaking, the BASC–2 would be more complicated to integrate successfully and effectively into CHILD's system than the short, straightforward PSC or SDQ. It has the potential to yield any number of data across a variety of psycho-social terrain on each student, and, especially if the BESS were adopted as well, would require additional capacity in order to absorb the many moving parts of this measure. At this point, as much potential as the BASC–2 has, it is not clear whether it is worth the trouble given its cost and logistical complexities. It should be noted as well that an update of the BASC–2, the BASC–3 is scheduled to be released in the near future. This presents a dilemma for CHILD as there will be little to no research on the BASC–3 (aside from what the test-maker conducts), but investing in the BASC–2 could be investing in an outdated product.

CBCL. The Child Behavior Checklist, or CBCL, is the last measure under consideration in this review. It is similar to the BASC–2 in that it is a part of a larger system of assessment known as the Achenbach System of Empirically Based Assessment, or ASEBA. According to the test-maker, the ASEBA is a "comprehensive evidence-based assessment system . . . [that] assesses competencies, strengths, adaptive functioning, and behavioral, emotional, and social problems" (Achenbach, 2015a), The CBCL technically refers to a specific test within the ASEBA system, the parent rater

form for children aged 6–18. The ASEBA system also has a teacher report form (TRF), youth self-report form (YSR), and, like the BASC–2, a brief screening form known as the Brief Problem Monitory, or BPM. For ease of explanation, and because much of the research tends to incorporate the TRF and the YSR into discussion of the CBCL, use of the term CBCL in this review will refer to all these three rating systems (the CBCL, TRF, and the YSR). The BPM will be treated as the BESS was treated in the review of the BASC–2, as a related measure to be used either in conjunction with or independently of the CBCL.

The CBCL, which appeared in five out of the six meta-reviews consulted for this literature review, has been a prominent tool in the mental health field since the 1990s (Deighton et al., 2014; Hunter et al., 1996). As a part of the ASEBA it can be used for myriad purposes: screening, documenting problems for use in referral servicing, intake assessment and treatment planning, tracking progress, determining whether goals are met, and evaluating outcomes (Achenbach, 2011). It is widely used in mental health services, schools, medical settings, child and family services, research, and other related fields (Williams, 2008). There are versions for students, parents, and teachers and scores can be combined or kept separate to provide useful clinical or mental health information on each student (Achenbach, 2015a). In length, the CBCL is longer than most of the measures described previously, with the exception of the BASC–2. It has between 105 and 120 items depending on the version (Deighton et al., 2014), and has an average completion time of 10–20 minutes (Payakachat et al., 2012; Williams, 2008). The BPM has between 18–19 items depending on the version, and takes one to two minutes to complete generally (Achenbach, 2015b). The student body population fits well within the age range

of both the CBCL and the BPM, although the YSR can only be used for students aged 11 years or older.

Ensuring alignment with CHILD's constructs of interest is another important area when exploring the appropriateness of a measure. The CBCL, in this case, covers a wide range of behavior, social, and emotional symptoms (Levitt et al., 2007; Payakachat et al., 2012; Williams, 2008) organized around constructs based on *DSM-IV-TR* diagnoses. Test items and results are organized around internalized and externalized behavior problems, including aggressive, hyperactive, noncompliant, and undercontrolled behavior (externalizing), and anxious, depressive, and overcontrolled behavior (internalizing) (Payakachat et al., 2012). The BPM is designed to link with the CBCL, and so its test results fall along the same dimensions (Achenbach, 2015b). In both the CBCL and the BPM, content scales can be summed to provide a single, "total problems" scale useful in the assessment of change across time (McClendon et al., 2011).

One distinct advantage the CBCL has over the other measures is that it "may be especially useful for measuring symptoms related to psychiatric comorbidities in children with ASDs" (Payakachat et al., 2012, p. 490), or autism spectrum disorders. There is a body of research emerging on the use of the CBCL in the assessment and treatment of ASD that is not present with the other measures described above. Moreover, this research goes beyond just the use of the instrument as a screening measure. For instance, Pandolfi, Magyar, and Dill (2012) found the CBCL was able to discriminate between youth with ASD and emotional and behavioral disorders versus those with just ASD. Others have demonstrated that the CBCL and TRF are useful in the identification of ASD (Mazefsky, Anderson, Conner, & Minshew, 2011; So et al., 2013). Given the presence of ASD in

CHILD's student body, as well as school district representatives' tendency to consider CHILD as a viable placement for students with ASD, a measure that accurately assesses this population, especially emotional and behavioral comorbidities could be particularly valuable. Alignment with CHILD's constructs of interest then is bolstered by this emerging research.

From a psychometrics perspective, the CBCL is one of the most widely researched and well-established measures in the field (Achenbach, 2011). It has moderate to high concurrent validity with the *DSM-IV-TR* clinical checklists and diagnoses (0.49–0.87 and 0.27–0.6, respectively) and with other measures such as the BASC–2 (0.46–0.89). Its reliability is sound, including an internal consistency as high as 0.95–0.97 across all versions, and its test-retest reliability (8–16 days) has ranged from 0.6–0.95 in the research (Deighton et al., 2014). The BPM, although much newer than the CBCL, has been found to have strong psychometric properties as well, including an internal consistency of 0.91 and an overall correlation with the CBCL of 0.95. Furthermore, research suggests the BPM is able to identify significantly higher emotional and behavioral problems amongst children who have been identified by their caregiver as having one of a wide variety of psychiatric disorders, including developmental disorders and ASD, relative to children who have not been identified as having these disorders (Piper, Gray, Raber, & Birkett, 2014).

As the BESS is designed to be used in tandem with the BASC–2 (Pearson Education, Inc., 2015a), so was the BPM created to work in line with the CBCL (Achenbach, 2015b). Specifically it says, "the BPM is designed to monitor children's responses to interventions (RTIs) over periods of days, weeks, or months. It is also

designed to monitor children's functioning during the course of services in special

education . . . " although, it adds, "ultimate outcomes of interventions and services should

be evaluated by comparing outcome scores on more comprehensive instruments with

initial scores on the same instruments, such as the CBCL/6–18, TRF, and YSR"

(Achenbach, 2015b, para. 1–2). While this is suggested or implied by the BASC–2

publisher, it does not specifically articulate this recommendation at any point, at least in

its online materials (Pearson Education, Inc.). The coinciding use of the CBCL, TRF,

YSR, and the BPM would serve CHILD's needs for robust outcome measurement while

tracking and producing data more regularly in a way that minimized burden on limited

school resources. Barriers to implementation of such a system, however, would include

its potential financial cost as well as inability on CHILD's part to effectively integrate the

complexity of such a system into its ranks. In other words, the usability of the CBCL and

BPM could be limiting factors to its adoption by CHILD if CHILD does not take steps to

effectively and fully accommodate the instruments from the beginning. Because the

CBCL fares well on the other two dimensions of Glover and Albers (2007) model,

appropriateness for its intended use and technical adequacy, CHILD may well want to

consider these points in greater detail. The constructs measured by the CBCL and BPM

seem to have a better alignment with the mental health constructs of interest than the

other measures—not only do they assess for a comprehensive battery of emotional,

social, and behavioral problems, but a growing body of evidence is demonstrating their

effectiveness in use with autism spectrum disorder. The relevance of the data obtained

using the CBCL and BPM is likely to be high.

One final aspect to consider regarding the CBCL and BPM is its financial cost to CHILD. For a computer scoring starter kit featuring 50 forms each of the CBCL, TRF, and YSR, but not the BPM, it is $475. For the computer scoring BPM starter kit, it is $230. Each pack of 50 forms for the CBCL, TRF, YSR, and all forms of the BPM costs $30, except the parent form of the BPM which would cost $25. This means to fully implement the CBCL, TRF, YSR, and BPM (assuming annual assessment and a student body of 50) it would cost CHILD $705 for the first year, and $175 for each subsequent year. This is cheaper than the BASC–2, but somewhat more than BERS–2 and comparable to the SSIS–RS, essentially about $90 for an annual CBCL administration and $85 for every school-wide BPM administration. Another option would be to use the web-based platform to administer and score the reports. From a usability standpoint, the web-based platform would simplify and streamline the testing process by centralizing data into one location, reducing paperwork and labor by providing remote access to parents, and more than likely increasing the response rates due to easier accessibility. Pricing on this varies greater due to the availability of bulk purchasing options—from $95–$2,500 depending on how many forms and reports CHILD wants to purchase—but would be approximately double the annual cost of the written forms. The ease and simplicity of this system may outweigh these costs however.

In summary, the CBCL has a number of advantages. It is part of a larger system, like the BASC–2, so could be used in conjunction with other tools for a variety of purposes, including screening, documenting problems for use in referral servicing, intake assessment and treatment planning, tracking progress, determining whether goals are met, and evaluating outcomes. It has good alignment with CHILD's constructs of interest: not

only does it measure strengths, adaptive functioning, and a broad range of behavioral, social, and emotional problems, a large body of research indicates that it is also useful in the assessment and treatment of ASD. An even larger body of research dating to the 1980s shows the CBCL has strong psychometric properties, and in general it is one of the most widely researched and well-established measures in the field. Finally, the availability of the BPM for use in conjunction with the CBCL can provide monitoring of progress frequently over time in between robust and thorough outcome measurement. The only short-comings relative to the other measures are that, unlike the PSC and SDQ, it is not free, and given its size and scope, its implementation may be logistically unfeasible. These elements will be discussed with CHILD during the Micro-level dialogue.

Stakeholder perceptions and opinions survey. The overarching objective of this evaluation stems from CHILD's desire to make claims to expertise in the area of mental health. However, "expertise" is, by definition, something that must be determined mutually by the organization and its external stakeholders, or consumers of its services. Therefore, CHILD needs some way of understanding the perceptions and opinions of these consumers, particularly as they pertain to CHILD's work in the area of mental health. In order to track this information in a manner that is both valid and reliable, CHILD needs to incorporate a data-tracking tool that can be used regularly and confidently by school administrators.

The Stakeholder Perceptions and Opinions Survey is like a customer satisfaction survey. According to Hayes (1998), such questionnaires are typically conducted in four phases: determining survey questions or items, selecting the response format or the

scale(s) of measurement to be used, writing the introduction, and determining which items specifically to include based on the content areas most relevant to the purpose of the survey. These will be used as guidelines in adapting the original parent and school district representative surveys for future use as a tool for tracking stakeholder (e.g., parents and school district representatives) perceptions and opinions.

In Hayes's (1998) view, the first step in developing a survey is creation of a pool of items based on the areas of interest to the organization and its stakeholders or consumers. Hayes provides five guidelines for this process. First, the specific items should be relevant to the interests of CHILD and their consumers ("consumer" is used interchangeably with "stakeholder" from this point forward). Second, the wording should be concise to avoid confusing the respondent. Some of the questions in the first version of the survey may have been too wordy and so may need to be revised and shortened. Third, the content should be specific and unambiguous. Fourth, a good item should only contain one question. These last two points speak to the importance of creating test items that clearly ask about one specific thought or idea. Finally, the items should not contain double negatives. In addition to these guidelines, the reliability and validity of the test should be taken into consideration. He notes that the reliability of a test tends to increase with the number of test items, and tends to decrease with the number of respondents who share similar perceptions and opinions. For instance, there was a positive skew amongst parents and school districts representatives in their opinions about CHILD. Future questions may need to be refined and oriented in a way that allows for a more standard distribution of answers, thus increasing the reliability of the response sets. Finally, the validity of a test and its items can be tested a number of ways, such as review by other

subject matter experts or by determining correlations between specific items and the constructs they are measuring. Determining validity is less precise than establishing reliability. For the surveys used in the present evaluation, content validity was checked by several subject-matter experts, who provided feedback about the validity of individual test items. This process was used again for the Stakeholder Perceptions and Opinions Survey.

The second phase of survey development, according to Hayes (1998), is determining the response formatting for each item. He provides two primary methods, checklists and Likert scales, the latter of which was used extensively in the original parent and district representative surveys. Adapting these surveys to the Stakeholder Perceptions and Opinions Survey, the Likert scale was retained on questions when possible. A simple yes/no format was used when necessary, and open-ended questions were also be considered in light of Hayes's suggestions for test item creation.

The third step in creating a survey is to write the introduction. This, according to Hayes (1998) should explain the purpose of the questionnaire, provide instructions for completing it, and explain how the data it generates will be used, in simple terms. When explaining the survey's purpose, it is important to consider the impact this has on the respondent. Too much information could impact or bias their responses; on the other hand, an explanation could help them connect with the meaning of the survey, thus increasing their chances of completing it. The instructions should explain the type of measurement scale that is being used, and how the respondent should fill out his or her answers (e.g., whether or not it is a Likert scale, and if it is, what type). Finally, Hayes recommends keeping the introduction brief.

The last step in survey construction is to select the final items that will be used in the questionnaire. Consideration should be given to the type of constructs that the survey is targeting, and to ensure that each item pertains to one of those constructs. For instance, CHILD's overall objective is to understand perceptions of their impact on student mental or social and emotional health. The construct "mental health" however is broad and variable. The results from the initial surveys showed that respondents were interested in the specific areas, behavior, coping skills, and self-regulation. Therefore, test items should be tied to or ask directly about these constructs.

This information was incorporated into the Stakeholder Perceptions and Opinions Survey, and is reflected in the Recommendations section, where a sample of the two surveys—a parent version and school district representative version—are included. Also reflected in the final recommendation is the outcome of conversations with CHILD's Leadership Team about the viability and logistics of implementing such a survey, discussed in greater detail below in the Micro-level evaluation section.

Long-term outcomes survey. The decision to incorporate a Long-Term Outcomes Survey into the data-tracking system is a result of several factors. Interest in this area was first expressed by members of CHILD's Leadership Team during the preliminary stages of the evaluation, when the evaluation question will still being developed. The Leadership Team was interested for its own sake in the long-term outcomes of the students that enroll at CHILD. As the evaluation question emerged, however, this information began to be seen as potential data points that could facilitate CHILD's claim to expertise. While not related to mental health directly, outcomes of this nature—for example, high school graduation rates, employment and housing status,

college enrollment, etc.—nonetheless provide an indirect indication of overall progress and success. The link between education, mental health, and these types of "long-term" outcomes is reflected in the literature. Freudenberg (2007), for instance, demonstrates that education and health outcomes—including mental health outcomes—are closely linked, with more education predicting better health outcomes. Fergusson and Woodward (2002) found that adolescents with depression were at a significantly increased risk for poorer educational outcomes and unemployment. The presence of psychiatric conditions in adults also significantly reduces their chances of gainful employment (Ettner, Frank, & Kessler, 1997), providing further argument for addressing mental health issues in an educational (and therefore pre-employment) setting. Finally, Johnson and Burr's (2010) research indicates that students with mental health issues are more likely to drop out of high school. Based on the meso-level surveys, parents and school district representatives expressed a strong interest in data of this nature. Parents also expressed an overwhelming willingness to provide this type of data.

The long-term outcome component of the data-tracking system is the most straightforward in terms of the development of the survey itself. Its purpose is ultimately to prove CHILD's value to external stakeholders. Its content, which was determined through the meso-level survey process, as well as through early dialogues with CHILD's Leadership Team, was a collaborative process utilizing input from a variety of sources. These processes resulted in a collection of specific "outcomes," such as high school graduation and employment rates, which can also be thought of as indicators of more general student progress and success. According to the Compassion Capital Fund National Resource Center (2010), broad outcomes such as these should be measured by

more specific and observable data points, otherwise known as indicators. Working from this framework, the "outcome" under consideration in the Long-term Outcome Survey is general student welfare, progress, and success (which, as noted above, correlates closely with mental health status). In order to track this outcome, concrete, measurable indicators needed to be established. Hence, test items on this survey were designed to inquire about easily reportable, clearly classifiable information such as college enrollment status. The survey, which is fairly simple in its design and thus relatively easy to complete, was labeled an "Outcome" survey however, for sake of clarity amongst respondents and other stakeholders. CHILD's capacity to effectively implement this survey, as important a topic as the content of the survey itself, was discussed with the Leadership Team and is explored in greater detail in the next section on micro-level evaluation.

Micro-Level Evaluation: Dialogue With CHILD

This portion of the evaluation was conducted following the macro-level literature review described in the previous section. While results from the surveys suggested that CHILD should implement three different data-tracking instruments, questions remained about CHILD's capacity to incorporate these instruments into their day-to-day functioning in an effective manner. CHILD's Leadership Team was consulted and a dialogue ensued focusing on two areas: the availability of resources in both time and money to implement the data-tracking system, and the logistics of implementation. Each of the three components of the system was reviewed in terms of these two areas.

The most important component of the proposed data-tracking system is the mental health outcome measure that CHILD will use to track student progress while enrolled at CHILD. A summary of the literature review results described above was provided to

CHILD, including an overview of the six measures included in the final stage of the analysis. The BERS–2 was presented as an outcome measure intended for use in schools. Its advantage as a brief and affordable instrument was highlighted alongside its limitations as a strengths-based only measure. The SSIS–RS was praised for its comprehensiveness, relatively good alignment with CHILD's constructs of interest, and population fit. On the other hand, it is a more expensive instrument, takes longer to administer, and requires a mental health specialist to score and interpret the data. Two measures, the PSC and the SDQ were discussed as potentially adequate measures with one distinct advantage above the others: both of them would be free to CHILD. They are also brief, straightforward, and easy to administer, score, and interpret (e.g., do not require a mental health specialist). However, due perhaps to the limited size and scope of these instruments, they do not measure as concisely and clearly the mental health constructs that are of interest to CHILD's stakeholders.

The remaining two instruments in the review, the BASC–2 and the CBCL, can be grouped together as comprehensive assessment systems designed for a wide variety of uses in a wide variety of settings. Both have long and short forms that can be used in tandem with one another as comprehensive outcome measures and more routine progress monitors. Due to their size and scope, they are likely to provide the best coverage of the mental health constructs most relevant to CHILD and its stakeholders. At the same time, they are longer and possibly more cumbersome to administer across an entire student body population. In terms of cost, the CBCL is cheaper than the BASC–2, on par with the SSIS–RS. The CBCL also has a more robust literature base, and has been shown repeatedly to be effective in use with autism spectrum disorder. Both have a web-based

administration and scoring platform, theoretically providing easier access to its array of tools for CHILD's staff; the exception to this is the abbreviated measure corresponding to the BASC–2, the BESS, which does not have a web-based version—it is hand- and software-based only.

After this summary was presented to CHILD's Leadership Team, including the specific pricing of each instrument, CHILD made clear they were interested in the best instrument regardless of cost. A final analysis to determine the "best" instrument relies on the guidelines put forth by Glover and Albers (2007) described earlier, and is located in the Recommendations section below.

The other two components to the data-tracking system, the Stakeholder Perceptions and Opinions Survey and Long-term Outcome Survey were also discussed with CHILD's Leadership Team. Because these will be developed as a part of this evaluation, they will be available to CHILD at no additional cost. They present only logistical challenges in terms of the time they will require of administrators and staff. First, and most importantly, CHILD expressed a desire to put both surveys to use, expressing a willingness to put the necessary time and energy into their implementation. It was decided that the Stakeholder Survey would be sent out to parents and school district representatives annually as a part of CHILD's ongoing School Improvement Plan. This is a Washington State requirement that all NPAs engage in formal, documented activities designed to improve various aspects of their programming and functioning. CHILD's Leadership Team indicated this survey could be used to show auditors that CHILD is actively soliciting feedback from its stakeholders in order to improve their school. It was also determined that the Long-term Outcome Survey would be distributed

to parents of former CHILD students using a database CHILD keeps active for fundraising purposes. The details of implementation of both the Stakeholder Perceptions and Opinions Survey and Long-term Outcome Survey are included in the Recommendations section.

The final, implied objective of the Micro-level dialogue with CHILD was to understand more completely the context within which the proposed data tracking would be put to use. CHILD has been undergoing some significant changes in the past year, including a move to a new building about 20 miles south. Not only did this impact the course of the school year (classes were temporarily held in an adjacent gym), but it had an effect on staff turnover as well: a number of staff did not make the move with CHILD to the new location. The move was completed by late 2014 however, and, according to CHILD's Leadership Team, the school has stabilized since then.

Of primary focus in this discussion was the status of CHILD's mental health department, particularly in terms of staffing and personnel. According to their Leadership Team, the clinical training arm, in place at the beginning of this evaluation, has been disbanded, if only temporarily. There is an interest in revisiting this aspect of the program in two to three years, but for the time being, there are no practicum or pre-internship students involved at CHILD. According to CHILD, the task of training clinical students only to have them leave after a year was too much of a burden. Instead, CHILD has increased the number of paid, full-time mental health specialists on staff. They currently employ one Licensed Mental Health Counselor (LMHC) and two LMHC Associates (LMHCAs). These staff members help fulfill students' required therapy hours and are used in a support role for teachers and other staff. They would likely be involved in a part

of the Mental Health Outcome component of the data-tracking system, helping in some capacity with the administration, scoring, and reporting of the tests. The other two surveys would more likely be managed by administrative staff. CHILD also noted that they are trying to move to a younger student body population. However, it was reported that school districts are more likely to give younger students an aide rather than relocate them to a new school entirely. At the same time, CHILD has managed to successfully reduce the number of aggressively violent and/or destructive students, enrolling instead those students that tend to "shut down" rather than "act out." Finally, CHILD stated that they try not to accept students whose parents are "not involved" in the education of their child.

For the sake of clarity, the remaining details of the Micro-level dialogue will be incorporated into the Recommendations below as integral aspects to the recommendations themselves. This will include specific staff and task assignments, the design and flow of each data-tracking component, and the overall implementation plan. These recommendations will be provided to CHILD as the final result of this program evaluation.

Results and Discussion: Summary

The core function of this program evaluation was organized according to Bronfenbrenner's ecological model of mental health. The first phase was a meso-level inquiry into the perceptions, opinions, and needs of select CHILD stakeholders: the parents of enrolled students and the school district representatives responsible for their placement and other related academic affairs. Two surveys, one for each stakeholder group, yielded data that was used to shape inquiry at the macro- and micro-levels. The

main findings were that these stakeholders tend to have a good opinion of CHILD, but would tend to regard them more highly with objective evidence of student progress and outcomes, mental health and otherwise. Based on this information, it was determined that data should be tracked in three areas: student mental health outcomes, stakeholder perceptions and opinions, and long-term outcome data (such as high school graduation rates and housing and employment status).

The next objective of the evaluation was a macro-level inquiry into the types of instruments and tools that would best serve the needs of CHILD and its stakeholders as identified by the meso-level survey. The most robust aspect of this phase was a detailed literature review of existing and viable mental health outcome instruments, culminating in a list of six possible tools. The literature was also reviewed in regards to the Stakeholder Perceptions and Opinions and Long-term Outcome Surveys, not to seek existing surveys, but to understand how best to design and implement new surveys.

The results of this macro-level inquiry were then relayed to CHILD's Leadership Team as a part of the micro-level inquiry, which essentially was a dialogue designed to work out the logistics and feasibility of implementing the proposed three-part data-tracking system. The results of this dialogue were used in conjunction with the meso- and macro-level inquiries to produce a final list of recommendations, presented in the following section.

Recommendations

The initial question to this program evaluation, developed vis-à-vis a series of

dialogues with CHILD's Leadership Team as well as a survey of internal stakeholders,

was "How does CHILD claim expertise, particularly in the area of mental health?" It was

determined that, while CHILD has considerable faith in its own capacity to impact

positive change in its students, they had no corroborating objective or empirical evidence,

or any system in place to collect such evidence. Therefore, the aim of this evaluation was

to determine, what evidence should be collected in order to help CHILD establish a claim

to expertise, especially in the domain of mental health. To help answer this question, two

surveys were developed for CHILD's external stakeholders, or the primary utilizers of its

services: the parents of enrolled students and the school district representatives

responsible for these students. Results from these surveys indicated that CHILD would be

best served by a data-tracking system with three components: one to measure mental

health outcomes, another to track long-term student outcomes, and finally a measure of

external stakeholder perceptions and opinions. The detailed description and

implementation plans for each comprise the final recommendations for this program

evaluation, and are discussed presently.

The Mental Health Outcome Measure

The most important and substantial component of the data-tracking system is the

Mental Health Outcome Measure. By implementing such a measure regularly and

appropriately, CHILD will begin to develop a database of information on the mental

health of its students. The primary purpose of this database will be to objectively inform

CHILD and its stakeholders about the progress these students are making while at

CHILD, so that reasonable inferences about their program might be made and claims to expertise eventually established. CHILD could also use the data on a student-to-student basis to improve programming or customize it to the varying needs of each individual.

Through a multi-stage review process, six measures were identified as potentially useful to these ends. An in-depth description and analysis of each of these measures was presented above using three primary guidelines put forth by Glover and Albers (2007): the appropriateness for its intended use, its technical adequacy, and its usability. For the most part, the six measures under consideration are all technically adequate. They have all been normed on nationally representative samples, and have been shown to have good reliability and validity. The psychometric profiles of each vary to some degree, but all could be described as sufficient or adequate based on the standards typically observed in the field. None have any psychometric properties, in other words, that would otherwise rule them out.

Furthermore, CHILD made clear that the cost of the measure should not factor in to the final decision. As such, the appropriateness for the intended use of the measures and their usability (aside from cost) become more relevant factors in the decision-making process. In Glover and Albers's (2007) view, a measure is appropriate for its intended use if it is compatible with the local service delivery needs, aligns with the constructs of interest, has sufficient theoretical and empirical support, and fits with the target population, and its usability is assessed across six domains: cost, feasibility, level of stakeholder "buy-in," available infrastructure for implementing the measure, the availability of special accommodations, and the relevance of the obtained data.

In terms of usability, the BERS–2 meets well with the needs of the data-tracking system. Its relatively short administration time (average completion time $=$ 10 min.) and design for use in schools make it a highly feasible measure, and CHILD's available infrastructure appears capable of adapting to and incorporating the measure without problem. Its appropriateness for the intended purpose also matches up well, with one exception. As a strengths-based measure, the BERS–2 does not overtly or explicitly measure the primary constructs of interest: behavior, self-regulation, and coping skills. While inferences could be made from the test results about a student's, or student body's level of mental health, the lack of alignment with the constructs of interest could decrease stakeholder "buy-in," and potentially soften any claims to expertise.

The SSIS–RS does not have this problem. In fact, its greatest asset is probably is its fit with the demographics of CHILD's student body as well as with the constructs of interest. In general, it is quite appropriate for the intended purpose of the measure. However, its usability is diminished by the fact that it requires a mental specialist to score and interpret the data. This places a potentially undue proportion of the burden on CHILD's small team of mental health professionals, whereas the protocols of the other instruments would effectively distribute the work load across the entire staff of teachers and program assistants. Furthermore, the level of specialized training could decrease the level of "buy-in" amongst staff if the data it produces is viewed as overly technical, jargonized, or otherwise inaccessible to the untrained observer.

The next two measures under consideration, the PSC and SDQ have a high degree of usability. They both are very straightforward, easily comprehendible, and brief in length. They do not require specialists for scoring and interpretation, and they are the

only two that are freely available to CHILD should they decide to use them. However, as already noted, cost is not a factor for CHILD, and so this advantage is neutralized.

In terms of appropriateness, the PSC has fair alignment with the constructs of interest, but it has several short-comings. It has a limited age range relative to the other measures, and does not have a version for teachers to complete, effectively reducing the amount of data relative to the other measures by a third. Finally, its brevity is convenient but limiting in terms of the breadth and depth of data it produces. The PSC was designed for use in physicians' waiting rooms as a screening measure to help physicians gain insight into their patients' mental health status. While it has been shown to be effective in tracking mental health progress, it may not be suitable in assessing the varying problems and needs of an entire student body, many of whom have been diagnosed with an autism spectrum disorder. There is no research indicating the PSC is useful in working with this population specifically.

The SDQ, which is similarly straightforward and easy to administer, has the added benefit of measuring multiple constructs across a wider age range. It appears to have been used more frequently in school than the PSC, and its presence in the literature is generally more substantial. However, it suffers the same shortcoming as the PSC in that it is not designed for use with autism spectrum disorder (ASD). In fact, its content scales align much more with the emotionally and behaviorally disturbed (EBD) population than with ASD. Given CHILD's interest in moving away from the EBD population, it would make little sense to implement a measure meant for this group.

The last two instruments can similarly be lumped together, albeit for different reasons. In many ways they are very different from the PSC and the SDQ, because they

are both a part of or are comprised of a comprehensive system of mental health assessment. Both the BASC–2 and the CBCL (and its family of instruments known as ASEBA) can be used for virtually any assessment need across virtually any setting. Both are ubiquitous in the literature, in part because they have been around for a considerable amount of time. Due to their scale and scope, they both have the potential to produce a considerable amount of data on each student, and across the school as a whole. This same characteristic could potentially be an obstacle to effective utilization, however. Both measures have over 100 items on each version, and take between 20–30 minutes to complete per administration (the CBCL is shorter by about 25–50 items, taking about 5–10 minutes less). The amount of data they produce on each student may also be somewhat beyond CHILD's needs, at least at first. Because the primary purpose of the data-tracking system is to track the overall mental health of CHILD's students, the breadth and depth of the instruments could potentially be extraneous.

Both the BASC–2 and the CBCL (ASEBA), as comprehensive systems, have addressed this problem however. In addition to their full-length protocols, they have also supplied abbreviated versions to be used in conjunction with the full versions. More in keeping with the brevity and simplicity of the PSC and the SDQ, the BESS (BASC–2) and the BPM (CBCL) are designed to be used as supplementary to their "parent" versions. The ASEBA website (Achenbach, 2015a) even suggests, in cases such as CHILD's, the use of the CBCL as a pre-/post- measure, with periodic use of the BPM in between the pre-test and the post-test to track mental health progress. Because the BESS and the BPM were designed as basically miniature versions of the BASC–2 and the CBCL, the data generated by both could be easily synthesized and applied to suit the

needs of the data-tracking system, as well as for any other purposes CHILD may see fit later on down the line. Given the availability of both long and short versions, from a usability standpoint gives the BASC–2 and the CBCL and a sizable advantage over the other measures. The only significant difference between the two, based on the information collected thus far, is that the full version of the BASC–2 is somewhat longer than the CBCL, taking on average about 5–10 minutes longer to complete. While not sizeable enough of a difference to rule it out, it's worth noting in the overall assessment of the instruments.

The last category of assessment, using Glover and Albers' (2007) guidelines, is the appropriateness of the measure for its purpose. As already stated, the BASC–2 and CBCL share many similarities, and it is, for the most part, no different in terms of their appropriateness for the needs of CHILD and the data-tracking system. They both draw data from three informant types (parent, teacher, and youth self-report), fit well with CHILD's age demographic, and have accumulated a considerable amount of theoretical and empirical support over time. Given their breadth and depth, they even align well with the constructs of interest, and would do much to inform CHILD about their students' mental health in terms of their self-regulation, behavior, and coping skills.

The one significant difference to emerge in a comparative review of the literature is that the CBCL (and ASEBA) works well with autism spectrum disorder (ASD). Research has demonstrated a level of effectiveness with this population that is not apparent with the BASC–2, or any of the other measures for that matter. In addition, the CBCL is by far the most researched instrument in relation to ASD, and its utility is well established. This is a clear advantage of the CBCL as it pertains to the proposed data-

tracking system, and ultimately what sets it apart from the BASC–2 and the others as the best candidate for implementation at CHILD. Therefore, the formal recommendation for the Mental Health Outcome Measure is the CBCL (ASEBA), to be used in conjunction with the BPM. The details of implementation should occur as follows:

The full version of the ASEBA, including the CBCL (Child Behavior Checklist), the TRF (Teacher Report Form), and the YSR (Youth Self-Report), should be administered twice for each student, one upon enrollment, and once again at the end of their enrollment, just prior to transitioning out of CHILD. This will provide CHILD with a detailed profile of each student's mental health, and how this changed over the course of their enrollment. This type of data was specifically requested by school district representatives. In the meantime, CHILD should implement, if possible, all three versions of the BPM (Brief Problem Monitor), including the teacher, parent, and student self-report forms on a regular basis, to generate more immediate data about student progress, and to monitor more closely their mental health status as they progress. This data garnered from this process will serve the dual purpose of demonstrating CHILD's impact on student mental health, while allowing CHILD to customize and adjust programming as needed to maximize the benefits for each student.

Based on conversations with CHILD's Leadership Team, it was determined that parents should receive a CBCL upon their child's enrollment at CHILD. Age-eligible students (11–18 years-old) should also receive a full-length YSR. Because teachers will be unfamiliar with the student upon enrollment, they will be asked to complete a full-length TRF by one month from the date of enrollment, or the first available progress report deadline, whichever is longer. The BPM will be administered quarterly to teachers

on schedule with the progress reports due in November, February, May, and August. It is

recommended that parents and students also complete these forms on a quarterly basis,

despite CHILD's concern of low engagement amongst these stakeholder groups. With

consistent administration of the BPM, participation may increase as they become more

familiar with the measure, and as data is presented back to them from the teacher

evaluations.

All test administrations, the CBCL, YSR, TRF, and BPM forms should be

completed using ASEBA's web-based service. This allows for all test administrations and

scoring to occur online. Data from each test is then compiled and made available to

CHILD and its staff. Using this service eliminates the need for producing, tracking, and

collecting paper-and-pencil forms, and reduces labor associated with data-collection and

entry. After initial startup costs, the annual fee for this service will be approximately

$350, depending on how many tests are purchased and administered. CHILD can receive

better bulk rates for purchasing more test administrations at once, referred to on the

ASEBA website as "e-units." The more e-units purchased at once, the less they are per

"unit." One e-unit is charged per administration and another again for scoring. Therefore

the range of cost for each test administration including scoring is $1.00-$1.50

(Achenbach, 2015a). The total annual cost then is based on at least four BPM

administrations to each student, one BPM administration each to teachers and parents, a

small number of full-length CBCLs, TRFs, and YSRs, and the annual subscription fee for

using the web-based service which is $50 (Achenbach, 2015a).

The Stakeholder Perceptions and Opinion Survey

The second component of the data-tracking system involves an annual survey of the primary utilizers of CHILD services: the parents of enrolled students, and the school district representatives with whom they regularly work. The purpose of this survey is to track objectively the perceptions and opinions these stakeholders have of CHILD and their services, particularly in the area of mental health. Because the survey administered for this evaluation was composed of many of the key elements this survey will need, it has been adapted for use on an annual basis and can be found in Appendix F.

As discussed during the micro-level phase of the evaluation, the survey will be administered as a part of CHILD's School Improvement Plan, a state-mandated function that requires CHILD to demonstrate ongoing efforts to improve their programming. Distributed annually, both school district representatives and parents will be asked either via email (as with district representatives) or on paper (as with parents), to complete their respective versions of the survey online at SurveyMonkey.com. A survey-specific URL will be provided to each respondent, allowing for easy access and improving the chances of good response rates. This is how the survey was administered for the present project, which yielded response rates over 70% for each stakeholder group. The web-based software at SurveyMonkey.com then compiles the data automatically and is readily available to CHILD's administrative staff for analysis and interpretation. CHILD's executive director will oversee the annual administration of this survey. Because CHILD already has an annual subscription to this service, there is no added cost to CHILD for implementing these surveys.

The Long-Term Outcome Survey

The final component of the data-tracking system, the Long-term Outcome Survey, is designed to provide CHILD data about its former students as they transition beyond their K–12 education. This type of data, to include high school graduation status, college enrollment status, and other metrics, was identified as highly relevant to CHILD and its stakeholders, and furthermore, likely to impact perceptions and opinions of CHILD. While it would not constitute mental health data per se, it would provide a more global indication of student progress and success and lend credibility, potentially, to CHILD's claims to expertise.

The Long-term Outcome Survey will be distributed in the same fashion as the Stakeholder Perceptions and Opinions Survey, using the web-based service SurveyMonkey. Parents of former students will receive notice of the survey in CHILD's annual fundraising mailings. A note included in this mailing will invite them to participate in the survey by logging into a survey-specific URL at SurveyMonkey.com. Data entered via this online portal will then be automatically stored and compiled, available to CHILD's administrative staff for analysis and interpretation. As with the Stakeholder Perceptions and Opinions Survey, CHILD's executive director will oversee the annual administration of the Long-term Outcome survey.

Conclusion

The initial evaluation question guiding the findings of this project was, "How does CHILD claim expertise, especially in the area of mental health?" As the evaluation unfolded, it became clear that CHILD and its internal stakeholders have a strong belief in the efficacy and effectiveness of their program. CHILD as an institution, however, has no objective evidence supporting this notion, compromising their ability to claim expertise. Because the idea of "expertise" is a mutually-defined concept—meaning CHILD's ability to make claims of this nature is dependent on the perceptions and opinions of its external stakeholders—it became necessary to understand the types of data most relevant to these stakeholders, and how this data might impact their perceptions and opinions of CHILD. Therefore, the evaluation focused specifically on the ways CHILD can and should measure effectiveness in the domain of mental health.

Through a three-tiered, ecological process of survey administration, literature review, and dialogue with CHILD's Leadership Team, this information was uncovered, and a system designed for the purpose of tracking mental health, and other outcome data. The final result was a series of recommendations for a three-part data-tracking system featuring a Mental Health Outcome Measure, a Stakeholder Perceptions and Opinions Survey, and a Long-term Outcome Survey. Using these tools, CHILD will be able to track the mental health of its students, the opinions (or "satisfaction") of its consumers, and the long-term outcomes of its former students. All three will serve as indicators of CHILD's impact, information that CHILD can use to make adjustments to their programming, or advertise as objective evidence of their "expertise."

The key for CHILD going forward is effective and consistent implementation of the proposed system. This necessitated a detailed plan for ongoing test administration, including the assignment of specific individuals to specific tasks, and the identification of specific dates for different parts of the system (e.g., BPM administration). If used correctly and consistently across time, this data-tracking system will yield a wealth of information likely to benefit CHILD's program as a whole, as well as the individual students within. As a long-term goal, CHILD is interested in creating a program that can be easily adopted in other areas, by other schools. Producing specific outcome data about its students would constitute a strong step in that direction.

References

Achenbach, T. (2011). CBCL. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 546–552). New York, NY: Springer.

Achenbach, T. (2015a). *Achenbach System of Empirically Based Assessment (ASEBA).* Retrieved from www.aseba.com

Achenbach, T. (2015b). *Brief Problem Monitor: Frequently asked questions.* Retrieved from www.aseba.com/bpmfaqs.html

Adelman, H. S., Taylor, L., Weist, M. D., Adelsheim, S., Freeman, B., Kapp, L., Lahti, M., & Mawn, D. (1999). Mental health in schools: A federal initiative. *Children's Services: Social Policy, Research, and Practice, 2*(2), 99–119. doi: 10.1207/s15326918cs0202_3

Bronfenbrenner, U. (1994). Ecological models of human development. In *International encyclopedia of education* (2nd ed., Vol. 3). Oxford, England: Elsevier. (Reprinted from *Readings on the development of children* (2nd ed.), pp. 37–43, by M. Gauvain & M. Cole (Eds.), 1993, New York, NY: Freeman)

Brown-Nagin, T. (2012). *Elementary and Secondary Education Act of 1965*. Retrieved from http://www.enotes.com/elementary-secondary-education-act-1965-reference/

Buckley, J. A., & Epstein, M. H. (2004). The Behavioral and Emotional Rating Scale-2 (BERS-2): Providing a comprehensive approach to strength-based assessment. *The California School Psychologist, 9*, 21–27. doi: 10.1007/BF03340904

Buckley, J. A., Ryser, G., Reid, R., & Epstein, M. H. (2006). Confirmatory factor analysis of the Behavioral and Emotional Rating Scale-2 (BERS–2) Parent and Youth Rating Scales. *Journal of Child and Family Studies, 15*(1), 27–37. doi: 10.1007/s10826-005-9000-2

Center for Mental Health in Schools. (n.d.). *Mental health in schools: An overview*. Retrieved from smhp.psych.ucla.edu/

Children's Institute for Learning Differences. (2014). *Children's Institute for Learning Differences*. Retrieved from www.childnow.org

Cohen, K. (2006). NCLB and the states. *PBS Newshour*. Retrieved from http://www.pbs.org/newshour/indepth_coverage/education/no_child/federalvstate.html

Collaborative for Academic, Social, and Emotional Learning. (2013). *What is SEL?* Retrieved from http://casel.org/why-it-matters/what-is-sel/

Compassion Capital Fund National Resource Center. (2010). Measuring outcomes. *Strengthening Nonprofits: A Capacity Builder's Library*. Retrieved from www.strengtheningnonprofits.org/resources/guidebooks/MeasuringOutcomes.pdf

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *Journal of Special Education, 47*(2), 71–82. doi: 10.1177/0022466911420877

Crosby, J. W. (2011). Test review [Review of the test *Social Skills Improvement System Rating Scales*, by F. M. Gresham & S. N. Elliott]. *Journal of Psychoeducational Assessment, 29*(3), 292–296. doi**:** 10.1177/0734282910385806

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

DiStefano, C., & Kamphaus, R. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders, 15*(2), 93–102. doi: 10.1177/10634266070150020401

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health,* 8(14). doi:10.1186/1753-2000-8-14

Doll, B., Spies, R., & Champion, A. (2012). Contributions of ecological school mental health services to students' academic success. *Journal of Educational and Psychological Consultation, 22*(1–2), 44–61. doi: 10.1080/10474412.2011.649642

Dowdy, E., Furlong, M., Eklund, K., Saeki, E., & Ritchey, K. (2010). Screening for mental health and wellness: Current school-based practices and emerging possibilities. In B. Doll, W. Pfohl, & J. S. Yoon (Eds.), *Handbook of youth prevention science.* New York, NY: Routledge.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432. doi: 10.1111/j.1467-8624.2010.01564.x

Education for All Handicapped Children Act, Pub. L. No. 94-142 (1975).

Elementary and Secondary Education Act, Pub. L. No. 89-10 (1965).

Elias, M. J. (2004). The connection between social-emotional learning and learning disabilities: Implications for intervention. *Learning Disability Quarterly*, *27*(1), 53–63. doi: 10.2307/1593632

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., Kessler, R., Schwab-Stone, M. E., & Shriver, T. P. (1997). *Promoting social and emotional learning: Guidelines for educators.* Alexandria, VA: Association for Supervision and Curriculum Development.

Ettner, S. L., Frank, R. G., & Kessler, R. C. (1997). The impact of psychiatric disorders on labor market outcomes. *ILR Review, 51*(1), 64–81. doi: 10.1177/001979399705100105

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. (2011). *Program evaluation: Alternative approaches and guidelines* (4th ed.). Boston, MA: Pearson.

Fergusson, D. M., & Woodward, L. J. (2002). Mental health, educational, and social role outcomes for adolescents with depression. *Archives of General Psychiatry, 59*(3), 225–231. doi:10.1001/archpsyc.59.3.225

Forness, S. R. (2005). The pursuit of evidence-based practice in special education for children with emotional or behavioral disorders. *Behavioral Disorders, 30*(4), 311-330. Retrieved from http://www.tartarus.ed.utah.edu/users/robert.o'neill/SPED%2051106110/Readings/forness05.pdf

Freudenberg, N., & Ruglis, J. (2007). Reframing school dropout as a public health issue. *Preventing Chronic Disease, 4*(4), 1-11. Retrieved from http://www.cdc.gov/pcd/issues/2007/oct/07_0063.htm

Gall, G., Pagano, M. E., Desmond, M. S., Perrin, J. M., & Murphy, J. M. (2000). Utility of psychosocial screening at a school-based health center. *Journal of School Health, 70*(7), 292–298. doi: 10.1111/j.1746-1561.2000.tb07254.x

Glover, T., & Albers, C. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135. doi: 10.1016/j.jsp.2006.05.005

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Quesionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1337–1345. doi: 10.1097/00004583-200111000-00015

Greenberg, M. T., Kusché, C. A., & Riggs, N. (2004). The PATHS curriculum: Theory and research on neurocognitive development and school success. In J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.), *Building academic success on*

social and emotional learning: What does the research say? (pp. 170–188). New York, NY: Teachers College Press.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*(6), 466–474. doi: 10.1037/0003-066x.58.6-7.466

Greene, R. W. (1998). *The explosive child: A new approach for understanding and parenting easily frustrated, chronically inflexible children*. New York, NY: HarperCollins.

Greene, R. W. (2011). Collaborative Problem Solving can transform school discipline. *Phi Delta Kappan, 93*(2), 25–28. doi: 10.1177/003172171109300206

Greene, R. W., Ablon, J. S., & Martin, A. (2006). Innovations: Child psychiatry: Use of Collaborative Problem Solving to reduce seclusion and restraint in child and adolescent inpatient units. *Psychiatric Services, 57*(5), 610–616. doi: 10.1176/appi.ps.57.5.610

Greene, R. W., Ablon, J. S., Monuteaux, M., Goring, J., Henin, A., Raezer, L., Edwards, G., Markey, J., & Rabbitt, S. (2004). Effectiveness of Collaborative Problem Solving in affectively dysregulated youth with oppositional defiant disorder: Initial findings. *Journal of Consulting and Clinical Psychology, 72*, 1157–1164. doi: 10.1037/0022-006x.72.6.1157

Gresham, F., & Elliott, S. N. (2008). *Social skills improvement system (SSIS) rating scales*. Bloomington, MN: Pearson Assessments.

Hayes, B. (1998). *Measuring customer satisfaction: Survey design, use, and statistical analysis methods*. Milwaukee, Wisconsin: American Society for Quality.

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Educational and Psychological Measurement, 71*(4), 617–637. doi: 10.1177/0013164410382896

Hunter, J., Higginson, I., & Garralda, E. (1996). Systematic literature review: Outcome measures for child and adolescent mental health services. *Journal of Public Health Medicine, 18*, 197–206. doi: 10.1093/oxfordjournals.pubmed.a024480

Individuals With Disability Education Act, 20 U.S.C. § 1400 (2004).

Jellinek, M. S., Murphy, J. M., Robinson, J., Feins, A., Lamb, S., & Fenton, T. (1988). The Pediatric Symptom Checklist: Screening school-age children for psychosocial

dysfunction. *Journal of Pediatrics, 112*(2), 201–209. doi: 10.1016/s0022-3476(88)80056-8

Johnson, P., & Burr, J. (2010). Improving post-school outcomes for youth with mental health issues. *Secondary Transition State Planning Institute*. Retrieved from www.ode.state.or.us/data/schoolanddistrict/funding/sped/secc/ed_summary_npso 1-(2).ppt

Kamphaus, R., Thorpe, J., Winsor, A., Kroncke, A., Dowdy, E., & VanDeventer, M. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 6*, 1–15. doi: 10.1177/00131644070670021001

Kazdin, A.E. (2005). Evidence-based assessment for children and adolescents: Issues in measurement development and clinical application. *Journal of Clinical Child and Adolescent Psychology, 34*, 548–558. doi: 10.1207/s15374424jccp3403_10

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*, 163–191. doi: 10.1016/j.jsp.2006.11.005

Mahan, S., & Matson, J. L. (2011). Children and adolescents with autism spectrum disorders compared to typically developing controls on the Behavioral Assessment System for Children (BASC–2). *Research in Autism Spectrum Disorders, 5*(1), 119–125. doi: 10.1016/j.rasd.2010.02.007

Martin, A., Krieg, H., Esposito, F., Stubbe, D., & Cardona, L. (2008). Reduction of restraint and seclusion through Collaborative Problem Solving: A five-year, prospective inpatient study. *Psychiatric Services, 59*(12), 1406–1412. doi: 10.1176/appi.ps.59.12.1406

Mazefsky, C. A., Anderson, R., Conner, C. M., & Minshew, N. (2011). Child Behavior Checklist scores for school-aged children with autism: Preliminary evidence of patterns suggesting the need for referral. *Journal of Psychopathology and Behavioral Assessment, 33*(1), 31–37. doi: 10.1007/s10862-010-9198-1

McClendon, D. T., Warren, J. S., Green, K. M., Burlingame, G. M., Eggett, D. L., & McClendon, R. J. (2011). Sensitivity to change of youth treatment outcome measures: A comparison of the CBCL, BASC–2, and Y–OQ. *Journal of Clinical Psychology, 67*, 111–125. doi: 10.1002/jclp.20746

McDuffie, K. A., & Scruggs, T. E. (2008). The contributions of qualitative research to discussions of evidence-based practice in special education. *Intervention in School and Clinic, 44*(2), 91–97. doi: 10.1177/1053451208321564

Mills v. Board of Education of the District of Columbia, F. Supp. 866 (D. DC 1972).

Navon, M., Nelson, D., Pagano, M., & Murphy, M. (2001). Use of the Pediatric Symptom Checklist in strategies to improve preventative behavioral health care. *Psychiatric Services, 52*(6), 800–804. doi: 10.1176/appi.ps.52.6.800

Nickerson, A. B., Brosof, A. M., & Shapiro, V. B. (2004). Predicting positive outcomes for students with emotional disturbance. *The California School Psychologist, 9*, 39–49. doi: 10.1007/bf03340906

OSEP Center on Positive Behavioral Interventions & Supports: Effective Schoolwide Interventions. (2013). *Positive behavioral supports and the law*. Retrieved from http://www.pbis.org/school/pbis_and_the_law/default.aspx

Pandolfi, V., Magyar, C. I., & Dill, C. A. (2012). An initial psychometric evaluation of the CBCL 6-18 in a sample of youth with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(1), 96–108. doi: 10.1016/j.rasd.2011.03.009

PARC v. Commonwealth of Pennsylvania, 343 F. Supp 418 (E.D. Pa. 1972).

Payakachat, N., Tilford, J. M., Kovacs, E., & Kuhlthau, K. (2012). Autism spectrum disorders: A review of measures for clinical, health services and cost-effectiveness applications. *Expert Review of Pharmacoeconomics & Outcomes Research, 12*(4), 485–503. doi: 10.1586/erp.12.29

Pearson Education, Inc. (2015a). *Behavioral Assessment System for Children, 2nd ed. (BASC-2)*. Retrieved from http://www.pearsonclinical.com/education/products/100000658/behavior-assessment-system-for-children-second-edition-BASC–2.html

Pearson Education, Inc. (2015b). *Social Skills Improvement System (SSIS) Rating Scales*. Retrieved from http://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssis-rating-scales.html#tab-details

Piper, B., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of Brief Problem Monitor, an abbreviated form of the Child Behavior Checklist. *Psychiatry and Clinical Neurosciences, 68*(10), 759–767. doi: 10.1111/pcn.12188

Positive Behavioral Intervention and Supports. (2015). *Positive Behavioral Intervention and Supports*. Retrieved from www.pbis.org

President's New Freedom Commission on Mental Health. (2003). *Achieving the promise: Transforming mental health care in America. Final report*. (U.S. DHHS Pub. No. SMA-03-3832). Rockville, MD: U.S. Department of Health and Human Services.

PRO-ED, Inc. (2012). *Behavioral and Emotional Rating Scale–Second Edition (BERS–2)*. Retrieved from http://www.proedinc.com/customer/productView.aspx?ID = 3430

Renshaw, T., Eklund, K., Dowdy, E., Jimerson, S., Hart, S., Earhart, J., & Jones, C. (2009). Examining the relationship between scores on the Behavioral and Emotional Screening System and student academic, behavioral, and engagement outcomes: An investigation of concurrent validity in elementary school. *California School Psychologist, 14*, 81–88. doi: 10.1007/bf03340953

Reynolds, C., Kamphaus, R., & Vannest, K. (2011). Behavior Assessment System for Children (BASC). In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 366–371). New York, NY: Springer.

Robinson, K. E., & Rapport, L. J. (2002). Outcomes of a school-based mental health program for youth with serious emotional disorders. *Psychology in the Schools, 39*(6), 661–675. doi: 10.1002/pits.10064

Simon, J. A. (2005). The rights of individuals with dyslexia and other disabilities under the law. In J. R. Birsh (Ed.), *Multisensory teaching of basic language skills* (2nd ed.). Retrieved from http://textbooks.brookespublishing.com/birsh/authors.htm

So, P., Greaves-Lord, K., van der Ende, J., Verhulst, F. C., Rescorla, L., & de Nijs, P. (2013). Using the Child Behavior Checklist and the Teacher's Report Form for identification of children with autism spectrum disorders. *Autism, 17*(5), 595–607. doi: 10.1177/1362361312448855

Sugai, G., & Simonsen, B. (2012). Positive behavioral interventions and supports: History, defining features, and misconceptions. *Center for PBIS and Center for Positive Behavioral Interventions and Supports*. Retrieved from http://www.pbis.org/school/pbis_revisited.aspx

Volker, M. A., Lopata, C., Smerbeck, A. M., Knoll, V. A., Thomeer, M. L., Toomey, J. A., & Rodgers, J. D. (2010). BASC–2 PRS profiles for students with high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(2), 188–199. doi: 10.1007/s10803-009-0849-6

Walberg, H. J., Zins, J. E., Weissberg, R. P. (2004). Recommendations and conclusions: Implications for practice, training, research, and policy. In J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.), *Building academic success on social and emotional learning: What does the research say?* (pp. 209–217). New York, NY: Teachers College Press.

Wallbrown, A. R. (2013). *Concurrent and predictive validity of the Behavior and Emotional Screening System* (Master's thesis). Retrieved from http//:mds.marshall.edu/etd/587

Whilden, B. (2012). The elementary and secondary education act: A primer on reauthorization in 2010. *American Association of State Colleges and Universities.* Retrieved from http://www.congressweb.com/aascu/docfiles/ESEA%20PRIMER%20FINAL.pdf

Williams, S. T. (2008). Mental health screening and assessment tools for children. *The Northern California Training Academy*. Retrieved from www.humanservices.ucdavis.edu/academy

Wholey, J. S. (1996). Formative and summation evaluation: Responding to the challenges. *Evaluation Practice, 17*, 145–149. doi: 10.1016/s0886-1633(96)90019-7

youthinmind. (2012). *SDQ: Information for researchers and professionals about the Strengths & Difficulties Questionnaire*. Retrieved from www.sdqinfo.com

Appendix A

CHILD Logic Model

CHILD Program Logic Model
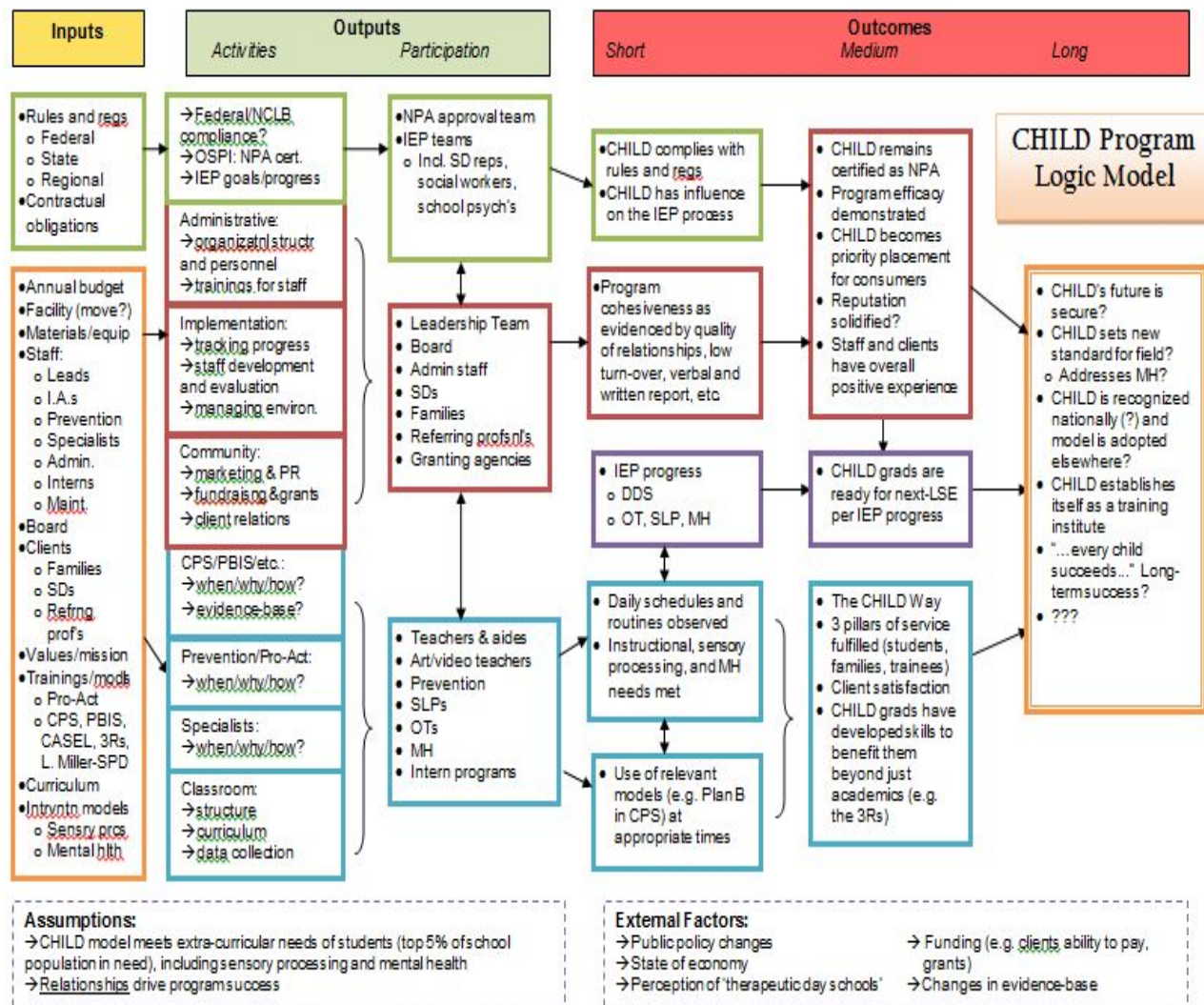
Appendix B

CHILD Program Evaluation Survey

1.) What is your general perception of CHILD, its services (school and clinic), and the impact CHILD has on clients and the community? Strengths? Areas needing improvement?

2.) What do you perceive as the purposes (goals, objectives) or guiding philosophy of CHILD?

3.) What do you believe is the theory or model CHILD follows? How does this lead to successful outcomes, and what aspects are most critical to success?

4.) What concerns do you have about the program services (school and clinic)? About outcomes, functions or operations? Any other issues?

5.) What do you hope to learn from the evaluation? What questions might you ask about CHILD and how might you use the information generated by such questions?

Appendix C

Content for Web-Based Survey-School District Representative Version

Children's Institute for Learning Differences (CHILD) is interested in learning more about the ways its programming impacts the mental health of its students. This survey is designed to help them gain a better understanding of their consumers' perceptions and expectations in this regard. As an important stakeholder and consumer of their services, your responses to this survey will provide them with valuable information, and may aid in their efforts to improve student mental health outcomes. There are 23 questions. No personal identifying information is required or stored (e.g., name, age, etc.). All information will be kept confidential and stored separately from the email address to which this survey was sent. Aggregated data will be presented to members of CHILD's Leadership Team, and may be used in program development, but none of your responses will be connected to any other information you've provided.

1. How long have you known about CHILD? [checks one of 5 boxes: 0-2 years; 3-5 years; 5-10 years; 10-15 years; 15+ years]

2. How would you describe your familiarity with CHILD? [3 choices: Not at all familiar; somewhat familiar; very familiar]

3. Have you known about CHILD outside of the context of your current position? [checks box, "Yes" or "No" ; if "No", respondent is skipped to Question #5]
    a. If you answered "Yes," please explain. [open-ended box, 100 words max.]

4. Please explain in what other context(s) you have known about CHILD (e.g. in a different professional role, as a parent, etc.)) [open-ended box, 100 words max.]

5. Approximately how many students have attended CHILD under your supervision? [checks one of 5 boxes: 0-1; 2-5; 6-10; 11-20; 21 or more; N/A]

6. How many of your district's students are currently enrolled at CHILD? [Multiple choice: 0; 1-2; 3-4; 5-7; 8+; I don't have access to this information]]

7. What are the most important factors when determining which non-public agency is the best program for a given student? Please rate the following criteria on a scale of 1-4, with 4 being Very/Always Important and 1 being Not at all/Never Important.
    a. Whether or not restraints/seclusions are used
    b. Whether the program is a good fit for the needs of the student
    c. Your understanding of the program's effectiveness with social and emotional intervention
    d. Your understanding of the program's effectiveness with academic intervention
    e. The likelihood the student will be able to return to public school within a given time frame
    f. Cost
    g. Other: _____

8. In regards to these same factors (as in Question #7), which aspects of CHILD appeal to you in your search for an appropriate program? Please rate the following criteria on a scale of 1-4, with 4 being "Very appealing" and 1 being "Not at all appealing/prohibitive."
   a. Their efforts to reduce/eliminate restraints, seclusions, and escorts
   b. The fit of their programming to the needs of the student
   c. Their program's effectiveness with social and emotional intervention
   d. Their program's effectiveness with academic intervention
   e. Their students' average length of stay
   f. Cost
   g. Other: _____

9. To the best of your knowledge, please rate whether you think CHILD has declined, remained about the same, or improved on the factors listed in Questions 7 and 8.
   a. Their efforts to reduce/eliminate restraints, seclusions, and escorts
   b. The fit of their programming to the needs of the student
   c. Their program's effectiveness with social and emotional intervention
   d. Their program's effectiveness with academic intervention
   e. Their students' average length of stay
   f. Cost
   g. Other: _____

10. Of the factors listed in Questions 7, 8, and 9, which would you like to see CHILD make the most improvements? Please rank them in order of importance, with 1 being the area of most importance to you.
    a. Their efforts to reduce/eliminate restraints, seclusions, and escorts
    b. The fit of their programming to the needs of the student
    c. Their program's effectiveness with social and emotional intervention
    d. Their program's effectiveness with academic intervention
    e. Their students' average length of stay
    f. Cost
    g. Other: _____

11. How likely are you to send a student to CHILD? [1 = Not at all likely; 2 = Somewhat unlikely; 3 = Somewhat likely; 4 = Very likely]

12. Using the same scale, how likely are you to send a student to CHILD who qualify under the following categories:
    a. Autism
    b. Behavioral/emotional disorder
    c. ADHD/Health Impaired
    d. Learning disability
    e. Per parent request
    f. Other: _____

13. How often do legal circumstances influence your decision to send a student to CHILD? [Never; Very little, Sometimes, Often]

14. CHILD claims that 100% of their students make "significant emotional, behavioral, and academic improvements." Please rate the degree to which you agree with this statement: [1 – Mostly disagree, 2- Somewhat disagree, 3 – Somewhat agree, 4 – Mostly agree]

15. How would you rate CHILD's ability to positively impact its students' social and emotional health? [Multiple choice: Very low; Somewhat low; Moderate; Somewhat high; Very high]

16. If CHILD was able to demonstrate its effectiveness using quantitative data other than IEP data, would this influence your decision to assign them students from your district (Yes or No) (Skip Logic to 18)

17. (Only if they answered 'Yes' to Q16). Please describe or give examples of the type of data that would influence your decision to assign CHILD students from your district, excluding IEP data.

18. If data was made available to you regarding CHILD's impact on its students' social and emotional development, what areas would be of most interest to you? Please check all that apply: [checks all that apply of the following: Self-regulation; Social interaction; Mood/attitude; Thought processes; Behavior; Coping skills; Problem-solving; Other: _____][checks all that apply; fills in blank on "Other" if checked]

19. If CHILD was able to demonstrate the long-term outcomes of its students, such as high school graduation rates or efficacy of intervention in domains such as social and emotional health, how would this impact your general opinion of CHILD? [1 = It wouldn't; 2 = A little; 3 = Somewhat; 4 = Considerably]

20. If CHILD were able to somehow demonstrate the long-term social and emotional outcomes of its students quantitatively, to what degree would this increase the likelihood you would send students to their school? [1 = It wouldn't; 2 = A little; 3 = Somewhat; 4 = Considerably]

21. If you have had or currently have students at CHILD, how would you rate your overall level of satisfaction with CHILD's services [Very low; Somewhat low; Moderate; Somewhat high; Very high]

22. If you do not currently have students at CHILD, please indicate why. Check all that apply:
    a. We often do, but are between student placements at this time

     b. Have I not heard of CHILD
     c. CHILD has not been proven effective with necessary intervention.
     d. CHILD costs too much given their level of efficacy.
     e. The school district has decided to send students to a more effective outside agency.
     f. The school district has decided to integrate students via special programs within the school district.
     g. Other: _____

23. The following is a list of approaches to social and emotional intervention in a school setting. Please place a check next to those you are familiar with:
     a. PBIS (Positive Behavior Intervention and Support)
     b. CASEL (Collaborative for Academic, Social, and Emotional Learning)
     c. CPS (Collaborative Problem Solving)
     d. The 3Rs (Regulation, Relationship, and Resilience)
     e. None of the above

24. If you are aware of other school-based programs targeting social and emotional health, please list those that have appeal to you. For each one that you list, please include a brief explanation as to why. If you don't know of any, you may skip this question. [5 open-ended blanks]

25. Please rank the following schools in order of overall preference:
     a. Overlake
     b. CHILD
     c. NWSOIL
     d. Renton Academy

26. How would you rate your level of satisfaction with CHILD's ability to address the social and emotional health of their students? [Very low; Somewhat low; Moderate; Somewhat high; Very high]

27. How would you rate your overall level of satisfaction with CHILD's services? [Very low; Somewhat low; Moderate; Somewhat high; Very high]

28. In your words, what is the most important contribution CHILD makes to the lives of your students? [open-ended box; 500 word limit]

Appendix D

Content for Web-Based Survey–Parent Version

Children's Institute for Learning Differences (CHILD) is interested in learning more about the ways its programming impacts the mental health of its students. To aid in the process, a survey is being distributed to the families of students currently enrolled at CHILD. The hope is that those closest to the students might offer their perceptions of and expectations for CHILD in this regard. As an important stakeholder and consumer of their services, your responses to this survey will provide them with valuable information, and may aid in their efforts to improve student mental health outcomes. There are 24 questions. No personal identifying information is required or stored (e.g., name, age, etc.). All information will be kept confidential and stored separately from the email address to which this survey was sent. Aggregated data will be presented to members of CHILD's Leadership Team, and may be used in program development, but none of your responses will be connected to any other information you've provided.

1. What is your relationship to the student you have enrolled at CHILD? [checks one of 4 boxes: Mother; Father; Legal Guardian; Other]

2. For approximately how long has your child been enrolled at CHILD? [checks one of 5 boxes: 0-6 months; 6 months-1 year; 1-2 years; 2-3 years; 3+ years]

3. Have you known about CHILD in any other capacity? [checks box, "Yes" or "No"]

4. If you answered "Yes" to the previous question, please explain. [open-ended box, 100 words max.]

5. Is your student enrolled through the school district or enrolled privately? [checks one of 2 boxes: "Enrolled through the school district;" or "Enrolled privately"]

6. How did you hear about CHILD? Please check one: [checks one of the following boxes: Public school official; IEP team; Teacher; Other professional (e.g. psychologist, social worker, occupational therapist, etc.); Another parent; Friend; Other (Please specify: _____)]

7. What did you perceive to be the greatest needs of your child when they first began at CHILD? Using numbers 1-5, please rank the following areas in order of highest or greatest need (1) to lowest or least amount of need (5): [uses numbers 1-5 to fill in 5 boxes next to the following areas: Learning/Academic Achievement; Sensory Processing; Mental Health; Problems of daily living (e.g. getting dressed, eating properly, personal hygiene, etc.); Speech and language]

8. How would you describe your investment in your child's mental health? [Checks one of 5 boxes: Not invested; Slightly invested; Moderately invested; Considerably invested; Extremely invested]

9. When your child first enrolled at CHILD, please describe your **general expectations** for their progress while at CHILD: [open-ended box; 250 word limit]

10. How would you describe your student's mental health needs when they first enrolled at CHILD? [checks one of 5 boxes, arranged left to right across the page: Very low/none; Somewhat low; Moderate, Somewhat high; Very high]
    a. After this question, respondents will be directed as follows: "If you answered "Very low/none," please skip to Question 16."

11. When your child first enrolled at CHILD, please describe your expectations for your child's progress in the area of **mental health** at CHILD. [open-ended box; 250 word limit]

12. Given what you knew about CHILD when your student first enrolled, how would you rate your expectations of CHILD's __intent__ to address your student's mental health needs? [checks one of 5 boxes, arranged left to right across the page: Very low/No expectations; Somewhat low; Moderate; Somewhat high; Very high]

13. Given what you knew about CHILD when your student first enrolled, how would you rate your expectations of CHILD's __ability__ to address your student's mental health needs? [checks one of 5 boxes, arranged left to right across the page: Very low/No expectations; Somewhat low; Moderate; Somewhat high; Very high]

14. How would you describe your student's __current__ mental health needs? [checks one of 5 boxes, arranged left to right across the page: Very low/none; Somewhat low; Moderate; Somewhat high; Very high]

15. *Given what you've learned about CHILD over time*, how would you rate your understanding of CHILD's __intent__ to address your student's mental health needs? [checks one of 5 boxes, arranged left to right across the page: Very low/none; Somewhat low; Moderate; Somewhat high; Very high]

16. *Given what you've learned about CHILD over time*, what do you believe is CHILD's __ability__ to address your student's mental health needs? [checks one of 5 boxes, arranged left to right across the page: Very low/none; Somewhat low; Moderate; Somewhat high; Very high]

17. Are you aware of the specific mental health services CHILD provides? [checks box, "Yes" or "No"]

18. Do you believe these services should be expanded or improved upon [checks box, "Yes" or "No"]

19. Please elaborate on the ways you believe these services could/should be expanded or improved upon. [Open-ended box - 250 word limit]

20. If information was made available to you regarding CHILD's impact on student mental health, what areas would be of most interest to you? Please check all that apply: [checks all that apply of the following: Self-regulation; Social interaction;

Mood/attitude; Thought processes; Behavior; Coping skills; Other: _____][checks all that apply; fills in blank on "Other" if checked]

21. The following is a list of approaches to mental health intervention in a school setting. Please place a check next to those you are familiar with: [a box is placed next to each of the following: "PBIS (Positive Behavior Intervention and Support)"; CASEL (Collaborative for Academic, Social, and Emotional Learning)"; CPS (Collaborative Problem Solving)"; The 3Rs (Regulation, Relationship, and Resilience)"]

22. If you are aware of other school-based mental health programs, please list those that have appeal to you. For each one that you list, please include a brief explanation as to why. [5 open-ended boxes; 100 word limit on each]

23. Please rate your level of satisfaction with CHILD's **mental health** services on a scale of 1-10, with 10 being the highest: [Open-ended, 2-digit box]

24. Please rate your level of **overall satisfaction** with CHILD's services on a scale of 1-10, with 10 being the highest: [Open-ended, 2-digit box]

25. CHILD is interested in knowing more about the long-term outcomes of its students, once they have transitioned beyond CHILD. Hypothetically speaking, if a program were initiated to track such outcomes, what is the likelihood you would be willing to participate?

26. Please check all of the types of information you might be willing to share in the interests of helping CHILD understand the long-term outcomes of its students: [checks all that apply of the following: "High school graduation/GED completion"; "College enrollment"; "College graduation"; "Employment status"; "Housing status"; "Other: _____"]

27. In your opinion, what is the most important contribution CHILD has made to the life of your child? [Open-ended box, 500 word limit]

Appendix E

CHILD Survey Report

Table of Contents

The following is a summary of results from surveys distributed to parents of students currently enrolled at CHILD and to school district representatives (referred to as 'SD reps' throughout) in April and May of 2014. The evaluator received 41 completed responses from parents and 31 completed responses from SD reps. Response rates and other statistics are included below. This summary compiles results from the two surveys, focusing especially on the data most relevant to the question, "How does CHILD claim its expertise?" Central to this question is a concern for improved outcome data collection, particularly as it pertains to stakeholder perceptions of CHILD's effectiveness. Both surveys use somewhat different wording to probe similar domains in this regard. For instance, the parent version focuses on student "mental health," while the school district version is concerned with student "social and emotional health." Each survey will first be reviewed independently and then compared across the questions they have in common.

Parent Survey

SURVEY RESPONDENTS

41 out of 71 (58%) completed surveys[3]
35 out of 48 (73%) households reporting
34 out of 46 (74%) students had at least one parent respond

PARENT INVESTMENT IN MENTAL HEALTH

Question: How would you describe your investment in your child's mental health?
- 96% of parents are "considerably" to "extremely invested"

PARENT EXPECTATIONS FOR PROGRESS

Question: Given what you've learned about CHILD over time, how would you rate CHILD's intent to address your child's mental health needs?
- 88% of parents believe CHILD intends to address student MH needs.

Question: Given what you knew about CHILD when your child first enrolled, how would you rate your expectations of CHILD's ability to address your child's mental health needs?
- 65% said "somewhat high" to "very high."

Question: Given what you've learned about CHILD over time, how would you rate CHILD's ability to address your child's mental health needs?
- 78% said "somewhat high" to "very high."

➔ The 13% increase in this area suggests that parents' belief in CHILD's ability to address student MH needs increases as they become more familiar with the school.

---

[3] Four additional surveys were partially completed; their responses were included in the final data set.

Question: When your child first enrolled at CHILD, please describe your general expectations for their progress while at CHILD. [Answers were sorted into themes and then listed below in order of salience. Bulleted quotes exemplify the type of response given under each theme.]

1. Social development
   - "That he would learn to cope better around other people outside of his family."
2. Improved behavior/Mental health
   - "Wanted her mental health to remain stable . . . But also wanted her challenged . . ."
3. Academic concerns
   - " . . . Improved academics."
4. Keeping child safe
   - "That he would be in an environment where he would be kept safe and that he would learn skills to be safe (for himself and others)"
5. Being accepted
   - "My expectations were that my son would become comfortable about himself and to receive help to uplift his self esteem [*sic*]."

PARENT PERCEPTION OF STUDENT NEEDS

Question: What did you perceive to be the greatest needs of your child when they first began at CHILD?
- #1: Mental health (17 out of 45 ranked it first)
- #2: Sensory processing (13 out of 45 ranked it first)
- #3: Learning/academic achievement (11 out of 45 ranked it first)

➔ Parents were also asked to rate their student's MH needs at intake (to CHILD) as well as their current MH needs. The results, which are compared in the graph below, suggest student MH needs decrease while enrolled at CHILD, at least according to the parents' perceptions.
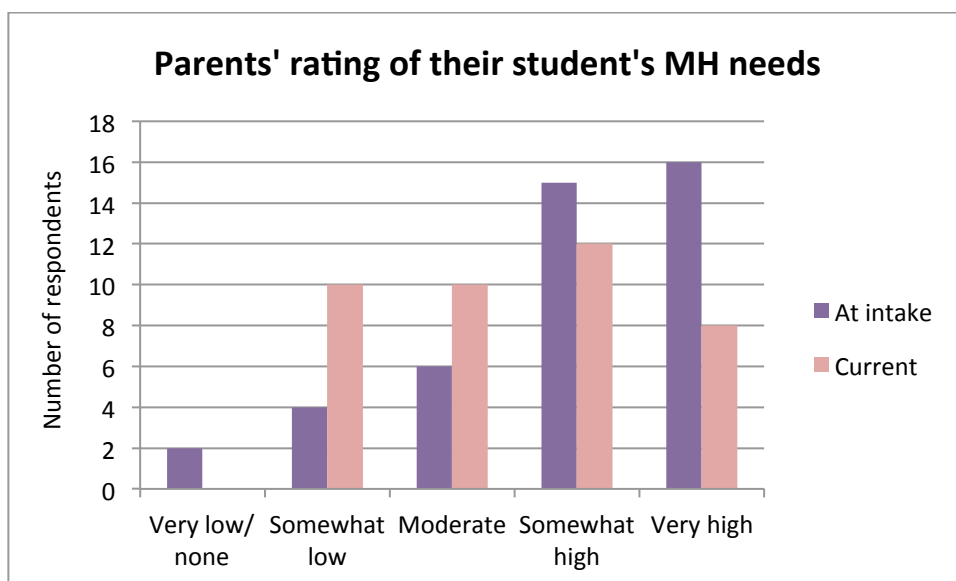
PARENT PERCEPTIONS OF CHILD SERVICES

➔ 79% of parents are aware of specific MH services at CHILD
➔ 70% believe these services should be expanded upon. When asked in what ways, specifically, should they be expanded, three themes emerged in order of salience:\

1. **More involvement on the part of the counselors**, both with the student, with the IEP team, and with the parents.
2. Reduced turnover amongst MH staff, **increased training or expertise in working with CHILD's student population**, and increased number of MH staff in general.
3. **Additional programming**, not only for students during school hours, but for students and their families outside of these hours. Several parents requested or alluded to more training for the students' caregivers, adding that, in the words of one respondent, "parents and loved ones should be carrying on what our children are learning while at school."
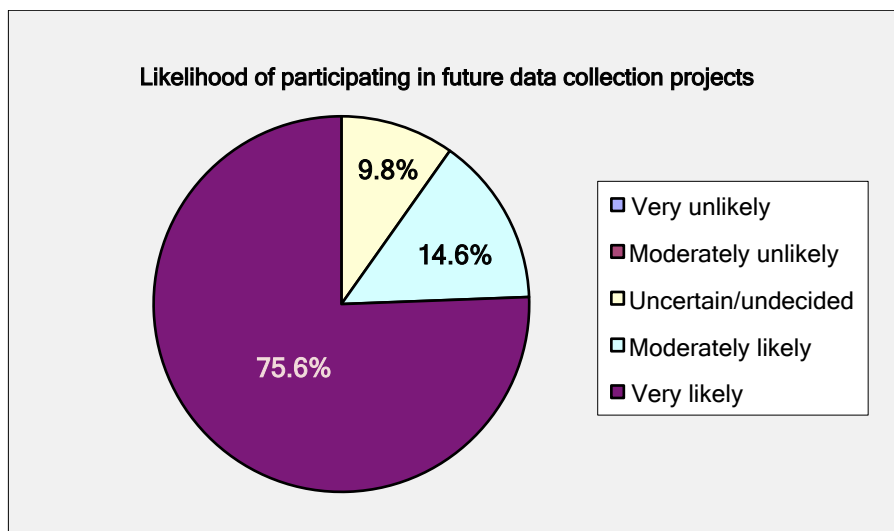
PARENT INTEREST IN AGENCY DATA COLLECTION

Question: If information was made available to you regarding CHILD's impact on student mental health, what areas would be of most interest to you?
   • Self-regulation (93%)
   • Coping skills and Behavior (85%)

Question: CHILD is interested in knowing more about the long-term outcomes of its students, once they have transitioned beyond CHILD. Hypothetically speaking, if a program were initiated to track such outcomes, what is the likelihood you would be willing to participate?
   • 0% responded "Moderately unlikely" or "Very unlikely"

Likelihood of participating in future data collection projects

9.8%

14.6%

75.6%

☐ Very unlikely
■ Moderately unlikely
☐ Uncertain/undecided
☐ Moderately likely
■ Very likely

Question: (In regards to the previous question,) what types of information would you be willing to share?

- At least 70% of the 41 respondents said they would be willing to share high school graduation/GED completion, employment status, college enrollment, college graduation and housing status.
- Other parent suggestions included social relationships, physical health, and hobbies. The sentiment in these responses was captured by this quote: "There isn't anything I wouldn't do to help promote the success we have found at CHILD."

<div align="center">

School District Survey

</div>

As noted above, "social and emotional" was used in place of "mental health" on this version of the survey due to the specific meaning "mental health" carries amongst educators. For the purposes of interpretation, however, the two phrases can be considered synonymous.

SURVEY RESPONDENTS

<div align="center">

31 out of 46 recipients have responded to the survey

67.4% overall response rate

24 out of 31 school districts had at least one responder

77.4% of all school districts represented

</div>

51.6% said they were "somewhat familiar" with CHILD, whereas 41.9% were "very familiar."
71% currently have at least one student enrolled at CHILD

SCHOOL DISTRICT PREFERENCES

Question: What are the most important factors when determining which NPA is the best program for a given student? [The following can be considered an aggregated list of the most important factors as understood by school district representatives, in order of importance.]

- Extremely/Always Important:
  -Whether the program is a good fit for the needs of the student
  -The program's effectiveness with social and emotional intervention
  -The program's effectiveness with academic intervention
- At least Somewhat/Occasionally Important, but *not* Extremely/Always Important:
  -The students' average or expected length of stay
  -The cost of the program
  -Whether or not restraints/seclusions are used
- Other factors listed by respondents:
  -Whether or not there are openings for the student (will they take a given student?)
  -Knowledge/expertise of staff

Question: How likely are you to send a student to CHILD?
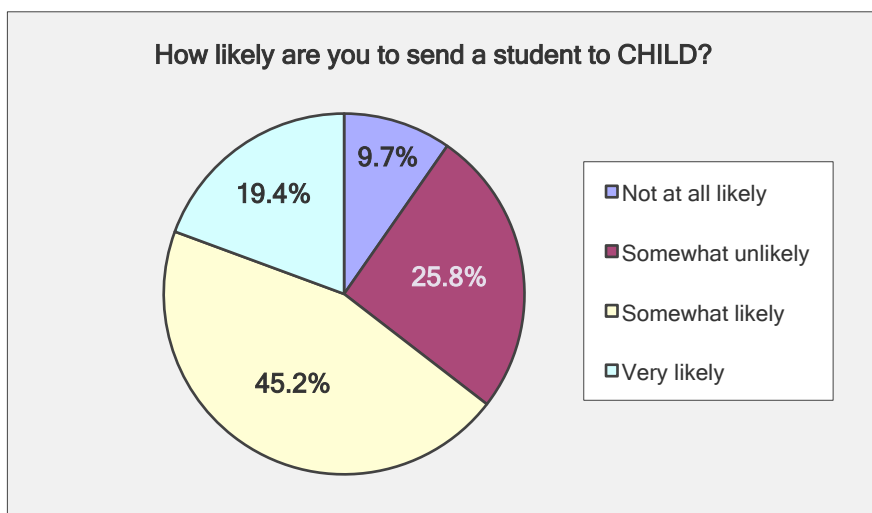
Question: How likely are you to send a student to CHILD who qualifies for special education services under the following categories? [listed in order of most to least likely, out of 31 total respondents]

- Autism
- Behavioral/emotional disorder
- Per parent request
- ADHD/Health impaired (14 out of 31 said "Not at all likely")
- Learning disability (22 out of 31 said "Not at all likely")
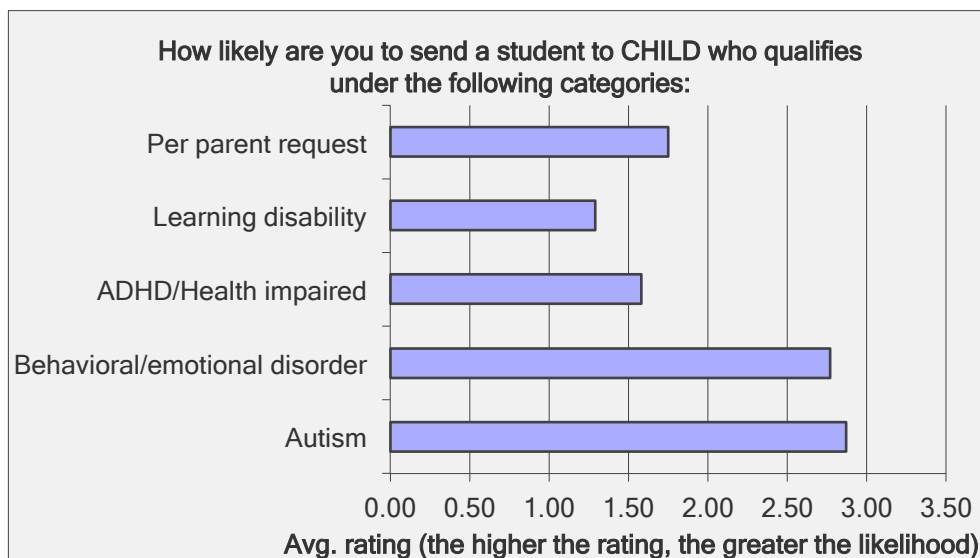
SCHOOL DISTRICT PERCEPTIONS OF CHILD

Question: Which aspects of CHILD appeal to you the most as you consider their program for a typical student? (Percentage of respondents who ranked the given aspect as either "Appealing" or "Very appealing" is included in parentheses)

- The fit of CHILD's programming to the needs of the student (71%)
- CHILD's efforts to reduce/eliminate restraints, seclusions, and escorts (65%)
- CHILD's effectiveness with social and emotional intervention (61%)
- CHILD's effectiveness with academic intervention (45%)
- Students' average length of stay at CHILD (42%)
- The cost of CHILD (34%)

Question: To what degree to you believe CHILD has improved, declined, or remained about the same on each of these separate dimensions?
- **About <u>half</u> of the respondents believe CHILD** has remained about the same along these dimensions, while **roughly <u>one-quarter</u>** believe they have made some improvements.

➔ Curiously, when asked in what areas they would like to see CHILD make the most improvements in, they ranked "fit of their programming to the needs of the student" first, with the program's "effectiveness with social and emotional intervention" coming in second. As these criteria were also ranked as two of the top three most appealing aspects of CHILD, these seemingly conflicting answers suggest that the question of how schools perceive the agency bears further investigation.
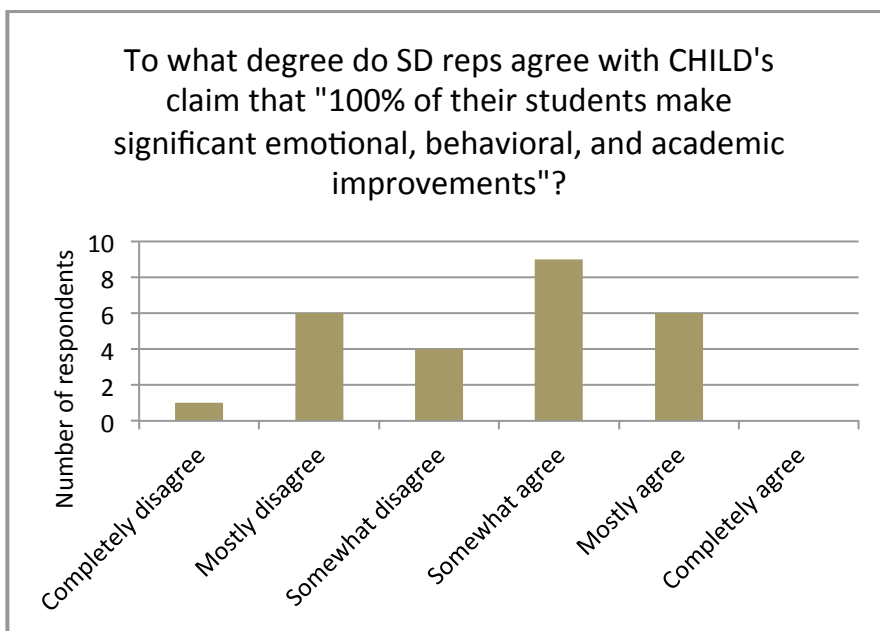
Questions regarding <u>data</u> and <u>future data collection</u> at CHILD:
- **71%** said that if "CHILD was able demonstrate its effectiveness using quantitative or empirical data other than IEP data," it would influence their decision to assign students to CHILD.
    - o When asked what types of data would influence their decision, respondents were most interested in some form of "behavioral" and/or "academic" data. Responses
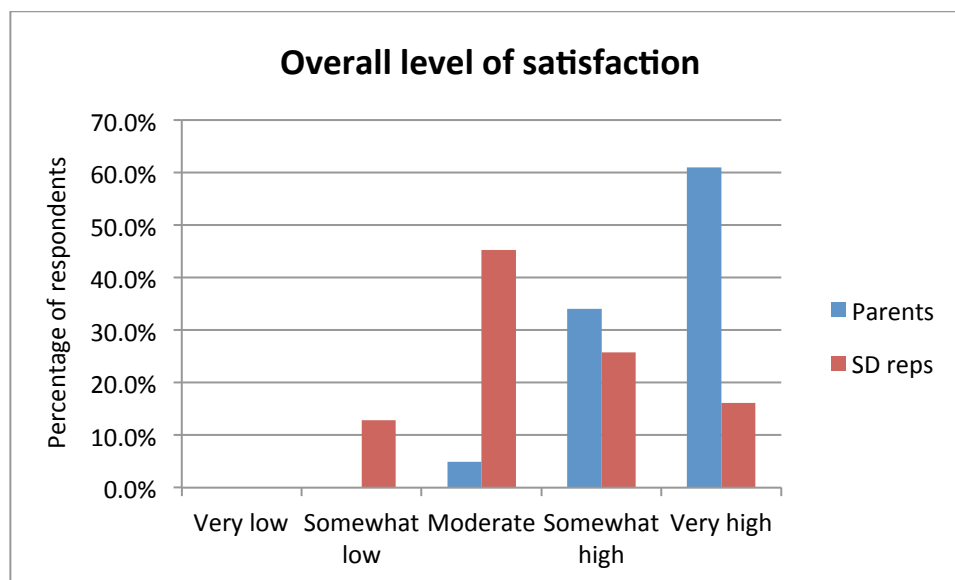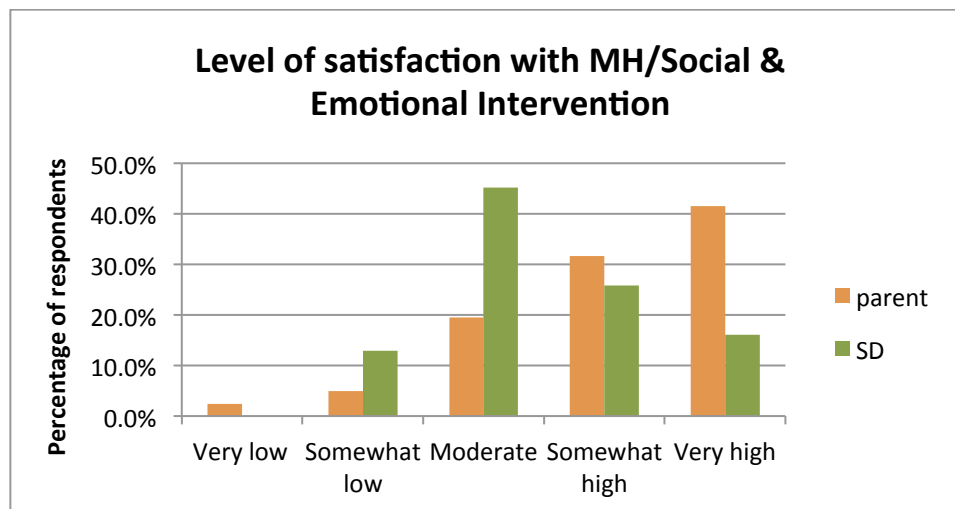
were more specific in regards to the types of behavioral data they would like to see. Salient examples include "disruption" to academics, functional/adaptive skills, and other social and emotional markers, particularly as collected in a pre/post-test format. Other data suggestions include average length of stay, percent of students graduating back to their home district, test scores, and attendance rates.

- If CHILD were able to provide data regarding their impact on **student social and emotional development**, the areas of most interest to the SD reps are, in order of interest*
  - Self-regulation
  - Behavior
  - Coping Skills
  - Social Interaction
  - Problem-solving
  - *Almost <u>no interest</u> was expressed in the categories "Mood/attitude" and "Thought processes"
- Demonstration of long-term outcomes for its students would "considerably" impact 65% of respondents' general opinion of CHILD. The remaining 35% would be "somewhat" impacted in their opinion of CHILD (so, **100% would be influenced by this type of data**).
- Similarly, 30 out of 31, or **97% would be more likely to assign students to CHILD if they were able to demonstrate *quantitatively* the long-term social and emotional outcomes of their students.**

Question: Please rank the following schools in terms of overall preference, with 1 being the highest, or most preferred (6 skipped this question due to insufficient knowledge of the programs)**:**

Ranking:	1	2	3	4	Rating Average
Renton Academy	3	6	12	4	2.68
The Overlake School	15	4	4	2	1.72
Children's Institute for Learning Differences (CHILD)	4	7	8	6	2.64
Northwest School of Innovative Learning (NW SOIL)	3	8	1	13	2.96

**Comparison Between
Parents & SD Reps**

Level of satisfaction with MH/Social & Emotional Intervention

Overall level of satisfaction

Question: In your opinion, what is the most important contribution CHILD has made to its students?

- Parents:
    - The **impact CHILD has on their students' self-esteem and sense of acceptance**. Some excerpts:
        - "He has rediscovered that he has worth," "he only associated school with being in trouble. Now he absolutely identifies himself as a student

and a learner," "she has learned that she can be successful in the classroom," and "he is not alone with his struggles."
- **Staff's understanding, encouragement, and advocacy.** Some excerpts:
  - "Our son feels love and devotion from his teachers," "[the teachers are] able to deal with behavior issues in a positive manner," and "he isn't worried about how staff will treat him on a daily basis."
- Improves students' desire to attend/participate in school

- SD reps:
  - **The staff**, particularly their "investment" in student progress, "level of caring," "collaborative" efforts with the school districts, and "respectful attitude" towards and "willingness" to help with challenging students.
    - In the words of one respondent, "at CHILD, the staff are not shocked by the behaviors and understand that it is part of the student's disability and why they are there to help."
  - Other prominent themes, in order of salience:
    - CHILD's **"willingness to take on tough cases,"** their impact on student social and emotional health, ability to tailor programming to the "unique needs of the 'whole' student," their focus on family integration, and their "variety of services."

<div align="center">

Summary

</div>

Perception and opinion of CHILD is more positive than negative. Both parents and school districts cite CHILD's staff as a prominent strength. Parents tend to value CHILD for their impact on their students' self-esteem and sense of belonging, while SD reps tend to appreciate CHILD's willingness to take on "tough cases." In general, responses indicate that SDs are more likely to send students to CHILD who qualify for services under the Autism and Emotional Disturbance categories than the Other Health Impairment (e.g., ADHD) and Specific Learning Disability Categories. Parents have a high opinion of CHILD while SD reps have a good, but more moderate opinion of CHILD. Parents would like to see more training for both the school staff and caretakers/families of the students, and would prefer more involvement on the part of the MH counselors.

In terms of overall satisfaction, parents rate CHILD higher than SD reps. While SD reps appreciate CHILD's ability to address the unique social and emotional needs of its students, they would like to see more progress in this area. In particular, there are some indications that SD reps are uncertain about the effectiveness of CHILD's programming, and would be interested in objective evidence demonstrating otherwise.

Prospects for Future Data Collection

Parents indicated they would be "very likely" to share long-term outcome data of their children after transitioning out of CHILD. Examples of data they'd be willing to share include: high school graduation/GED completion, employment status, college enrollment and graduation, and housing status.

Significantly, SD reps were nearly unanimous in their opinion that quantitative and/or long-term outcome data would influence their decision to send students to CHILD. Types of data in which SD reps may be interested:
- Behavioral, particularly through the use of pre/post measurements
    -Disruptiveness/aggression/incidences
    -Social behavior and interactions
    -Self-regulation/coping skills
    -Adaptive/functional
- Academic
    -Progress on assigned curriculum
    -Test scores
    -Ability to remain in the classroom
- Other:
    -Attendance rates
    -Length of stay
    -Graduation/exit rates
    -Aggregated population data

The survey findings demonstrate a desire for more objective data across a number of dimensions, especially among school district representatives. While it is not clear if CHILD's current lack of objective data is damaging to their reputation amongst parents and school districts, it is apparent that increased data collection in certain areas would be likely to improve consumer opinion of CHILD. However, CHILD's ability to collect this data is complicated by a number of factors including cost, time, and other issues such as differing privacy practices across school districts. More information is needed to understand CHILD's capacity for obtaining, storing, and utilizing this type of data. In the meantime, CHILD's efforts to develop their relationships with parents and the school districts in the surrounding areas may be working to offset some of the misconceptions and/or concerns SD reps have about CHILD's programming and the impact it has on students.

Appendix F

Stakeholder Perception and Opinion Survey

School District Representative Version

Children's Institute for Learning Differences (CHILD) is interested in understanding the perceptions and opinions others have of their programming and the impact it has on its students. The school district representatives with whom CHILD regularly works are an important stakeholder group in this regard. This survey, distributed annually, is part of a continual effort on CHILD's part to improve the quality of its services based on the needs of its students, their families, and the school districts that serve them. Please take a few moments to take the survey at [specific SurveyMonkey URL]. There are 20 multiple-choice questions which should take about five minutes to complete. No personal identifying information is required (e.g. name, age, etc.), and all information will be kept confidential and stored separately from the email address to which this survey was sent. Aggregated, de-identified data will be available to CHILD's administrative staff, and may be used in program development. CHILD thanks you for your time and input, as the information you provide will serve to improve the quality of CHILD's programming for the benefit of its students.

1. How long have you known about CHILD? [0-2 years, 3-5 years, 5-10 years, 10-15 years, 15+ years]
2. How would you describe your familiarity with CHILD? [Very familiar, Somewhat familiar, Not at all familiar]
3. How often does your school district use CHILD? [Frequently, Occasionally, Rarely, Not at all]
4. How likely are you to send a student to CHILD? [Very likely, Somewhat likely, Somewhat unlikely, Not at all likely]
5. To the best of your knowledge, please rate whether you think CHILD's program has improved, declined or remained about the same over the last year. [Improved, Remained about the same, Declined, I don't have enough information to answer accurately]
6. Based on what you know about CHILD, how would you rate their programming overall? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
7. Based on what you know about CHILD, how would you rate the impact of their programming on student social and emotional health? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
8. Does CHILD make available data on student social and emotional health? [If No, please skip to Question 10]
9. How does this data influence your opinion of CHILD? (Positively, Negatively, It does not)
10. Does this data make you more or less likely to consider CHILD as a placement option? (More, Less, It does not have an impact)
11. Does CHILD make available data on the long-term outcomes of its former students, such as high school graduation rates, college enrollment status, employment status, or housing status? [If No, please skip to Question 14]
12. To what degree do you believe these outcomes might be attributable to CHILD? [Very much so, Somewhat, A little, None at all]

13. How does this data influence your opinion of CHILD? [Positively, Negatively, It does not]
14. Does this data make you more or less likely to consider CHILD as a placement option? [More, Less, It does not have an impact]
15. Please rank the following schools in order of overall preference:
    a. Overlake
    b. CHILD
    c. NWSOIL
    d. Renton Academy
16. Which of these schools would you consider <u>first</u> for a student qualifying under the Emotional Disturbance category? [Overlake; CHILD; NWSOIL; Renton Academy]
17. Which of these schools would you consider <u>first</u> for a student qualifying under the Autism category? [Overlake; CHILD; NWSOIL; Renton Academy]
18. Which of these schools would you consider <u>first</u> for a student qualifying under the Other Health Impairment category? [Overlake; CHILD; NWSOIL; Renton Academy]
19. How would you rate your level of satisfaction with CHILD's ability to address the social and emotional health of their students? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
20. How would you rate your overall level of satisfaction with CHILD's services? [Very high, Somewhat high, Moderate, Somewhat low, Very low]

Parent Version

Children's Institute for Learning Differences (CHILD) is interested in understanding the perceptions and opinions others have of their programming and the impact it has on its students. The school district representatives with whom CHILD regularly works are an important stakeholder group in this regard. This survey, distributed annually, is part of a continual effort on CHILD's part to improve the quality of its services based on the needs of its students, their families, and the school districts that serve them. Please take a few moments to take the survey at [specific SurveyMonkey URL]. There are 16 multiple-choice questions which should take less than five minutes to complete. No personal identifying information is required (e.g. name, age, etc.), and all information will be kept confidential and stored separately from the email address to which this survey was sent. Aggregated, de-identified data will be available to CHILD's administrative staff, and may be used in program development. CHILD thanks you for your time and input, as the information you provide will serve to improve the quality of CHILD's programming for the benefit of its students.

1. How long have you known about CHILD? [0-2 years, 3-5 years, 5-10 years, 10-15 years, 15+ years]
2. How long has your child been enrolled at CHILD? [0-1 year, 1-3 years, 3-5 years, 5+ years]
3. How would you describe your familiarity with CHILD? [Very familiar, Somewhat familiar, Not at all familiar]
4. How likely are you to recommend CHILD to another parent? [Very likely, Somewhat likely, Somewhat unlikely, Not at all likely]
5. To the best of your knowledge, please rate whether you think CHILD's program has improved, declined or remained about the same over the last year. [Improved, Remained about the same, Declined, I don't have enough information to answer accurately]
6. Based on what you know about CHILD, how would you rate their programming overall? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
7. Based on what you know about CHILD, how would you rate the impact of their programming on mental health? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
8. Does CHILD make available data on student mental health? [If No, please skip to Question 10]
9. How does this data influence your opinion of CHILD? (Positively, Negatively, It does not)
10. Does this data make you more or less likely to recommend CHILD as a placement option to other parents? (More, Less, It does not have an impact)
11. Does CHILD make available data on the long-term outcomes of its former students, such as high school graduation rates, college enrollment status, employment status, or housing status? [If No, please skip to Question 14]
12. To what degree do you believe these outcomes might be attributable to CHILD? [Very much so, Somewhat, A little, None at all]

13. How does this data influence your opinion of CHILD? [Positively, Negatively, It does not]
14. Does this data make you more or less likely to recommend CHILD to other parents? [More, Less, It does not have an impact]
15. How would you rate your level of satisfaction with CHILD's ability to address the mental health of its students? [Very high, Somewhat high, Moderate, Somewhat low, Very low]
16. How would you rate your overall level of satisfaction with CHILD's services? [Very high, Somewhat high, Moderate, Somewhat low, Very low]

Appendix G

Long-term Outcome Survey

The Children's Institute for Learning Differences (CHILD) is interested in tracking the long-term outcomes of its former students such as high school graduation status, college enrollment status, employment status, and housing status. Please take a few moments to take the survey at [specific SurveyMonkey URL]. The survey has 12 questions and should take less than five minutes to complete. This information will be stored anonymously and used to help CHILD understand the impact it has on its students. In some cases, CHILD may publish compiled data (for instance, in their marketing materials) to help demonstrate this impact. However, the data you provide will in no way be linked publically or privately to any personally identifying information related to you or your family. Thank you for your time and effort and we hope you are doing well.

1. How old is your child currently?
2. For approximately how many years did your child attend the Children's Institute for Learning Differences (CHILD)? [0-1, 2-3, 4-5, 6+]
3. Approximately how old was your student when he/she transitioned out of CHILD? [8 or under, 9-11, 12-15, 16-18]
4. After transitioning out of CHILD, was your student ever re-enrolled in an independent placement, whether at CHILD or anywhere else? [Yes, No]
5. Did your child graduate from high school? [Yes, No, N/A]
6. Did your child enroll in any post-secondary technical or vocational programs? [Yes, No, N/A]
7. Did your child graduate from any post-secondary technical or vocational programs? [Yes, No, N/A]
8. Did your child enroll in any college coursework? [Yes, No, N/A]
9. Did your child graduate from college? [Yes, No, N/A]
10. If you answered 'Yes' to Question 9, please list the highest degree that your child has earned: [Associate's degree, Bachelor's degree, Master's degree, Doctoral or other professional degree]
11. What is your child's current living situation? [At home with parent/s, Independent with some supports (e.g. assistance from an aid or family member), Independent with no supports, Residential treatment, Group home, Other]
12. What is your child's employment status? [Full-time employee, part-time employee, self-employed (full-time), self-employed (part-time), Job training or transitional program, Not working, but seeking, Not working, not seeking]